Classical solutions of sigma models in (2+1) dimensions

Leese, Robert Anthony

CLASSICAL SOLUTIONS OF SIGMA MODELS IN (2+1) DIMENSIONS

by

ROBERT ANTHONY LEESE

A thesis presented for the degree
of Doctor of Philosophy at the
University of Durham

Department of Mathematical Sciences
University of Durham
Durham DH1 3LE
England.

March 1990

To Kathleen

Preface

This thesis is the result of work carried out in the Department of Mathematical Sciences at the University of Durham, between October 1987 and March 1990, under the supervision of Dr. R. S. Ward. No part of it has been previously submitted for any degree, either in this or any other university.

With the exceptions of chapters I and II, and appendix A, this work is believed to be entirely original. Chapters III and IV form the basis of papers which have been published in the *Journal of Mathematical Physics*[1] and *Physical Review D*[2] respectively. The material in chapter V is included in a paper to appear in *Nonlinearity*,[3] written jointly with Michel Peyrard and Wojciech Zakrzewski. However, this thesis contains only those portions of the paper that are due solely to the author. Finally, chapter VI has been written up[4] and submitted for publication in *Nuclear Physics B*.

I thank Richard Ward for his continued interest in my work, and for his careful supervision throughout; also Wojciech Zakrzewski for many interesting discussions arising out of his own work on the questions of stability and scattering in nonlinear σ-models. In addition, I have enjoyed stimulating conversations with Patrick Dorey and Ian Strachan. Financial support was provided by the Science and Engineering Research Council, whom I acknowledge with thanks.

CLASSICAL SOLUTIONS OF SIGMA MODELS IN (2+1) DIMENSIONS

by

ROBERT ANTHONY LEESE

Ph.D. Thesis, 1990

Abstract

This work is concerned with the large class of nonlinear scalar field theories known as σ-models, and in particular with their classical solutions. It is shown how the σ-models can admit solitons in (2+1) dimensions; and how, in many cases, these solitons can be classified topologically. For the Kähler σ-models, the instanton (*i.e.* static soliton) solutions are derived explicitly via the Bogomolny equations.

The main part of the thesis looks at the behaviour of solitons under the influence of small perturbations, and at their (classical) interactions. Attention is confined to the O(3) σ-model and its close relatives. A recurring theme is the ability of solitons to change in size as they evolve, a feature which is attributed to the conformal invariance of the theory.

There seem to be three possible approaches. In some special cases, the theory is integrable, in the sense that one can write down explicit time-dependent solutions. More often, one must resort to a numerical simulation, or else some sort of approximation. For theories that possess a topological lower bound on the energy, there is a useful approximation in which the kinetic energy is assumed to remain small.

All three of these approaches are used at various stages of the thesis. Chapter III deals with the properties of wave-like solitons in an integrable theory, and reveals some hitherto unseen behaviour. Chapters IV and V develop a numerical simulation based on topological arguments, which is then used in a study of soliton stability in the pure O(3) model. The conclusion is that the solitons are unstable to small perturbations, in the sense that their size is subject to large changes, even though their energy remains roughly constant. Chapter VI uses the slow-motion approximation to investigate soliton interactions in the O(3) model, and uncovers a plethora of possibilities.

Finally, some suggestions are made regarding possible directions for future research. In particular, attention is focussed on ways of modifying the O(3) model in an attempt to stabilize its solitons against changes in size.

Contents

Chapter I
Introduction

During the last 15 years there has been increasing awareness of the connection between many concepts in pure mathematics and theoretical physics. The physicist has found great benefit in learning a little pure mathematics; conversely, the mathematician has found new application for his own skills in helping to solve the problems of physics. Just one consequence of this fruitful exchange is the widespread interest in the large class of nonlinear field theories known as σ-models.

The aim of this thesis is to discuss the important properties of these models in general and then to concentrate on a few examples in particular. An effort is made to make explicit the relevance to both pure mathematics and theoretical physics, without emphasizing either aspect at the expense of the other. In fact, a large portion of the work uses numerical methods and various computing techniques; so, if one wants to attach any particular label to the overall approach, it should perhaps be that of applied mathematics.

We shall work mostly in (2+1) dimensions, in which an important feature of the σ-models is the existence of "solitons": roughly speaking, these are lumps (or sometimes waves) of energy that move on a two-dimensional plane at constant velocity and without changing shape. Two interesting problems are to look at how a soliton is affected by small perturbations and at what happens when two solitons collide. This thesis attempts to answer these questions for several σ-models, highlighting the different broad types of behaviour that can occur.

Before any detailed discussion, and to avoid confusion later on, it is worthwhile clearing up a small point of terminology: the word "soliton" was introduced in the 1960s by mathematicians to describe lumps of energy that were stable to perturbations and did not change either velocity or shape when colliding with one another. However, in the last 20 years all sorts of localized energy configurations have been called solitons in the literature. We shall go along with this looser definition. By a soliton we shall mean a lump of energy that moves at constant velocity and without changing shape; but we shall imply neither stability to perturbations nor a simple behaviour in collisions.

Solitons do not occur only in σ-models; in fact many partial differential equations have soliton solutions. However, it should be stressed that, for an equation picked at random, they are very much the exception rather than the rule. They can occur only when dispersion effects are exactly balanced by nonlinearities, for it is only then that a lump can move without changing shape. The simplest equation with this property is the wave equation, which is both linear and dispersionless; but in general the balance is much more delicate. In any case, the wave equation (as its name implies) has only wave-like solitons and not lump-like ones.

The major breakthrough in soliton theory was the discovery in the late 1960s of the so-called inverse scattering method,[5] which provides a recipe for writing down soliton solutions to a large number of equations. Since then several other ways of constructing solitons and looking at their interactions have been developed, and considerable progress has been made. However, most of it is confined to (1+1) dimensions. In (2+1) dimensions (and higher) the subject is not nearly so well understood, and many of the solitons known at present are simply extensions of familiar examples from (1+1).[6] One of the main tasks in soliton theory is to clarify the situation in higher dimensions.

An important feature of higher-dimensional theories is their topology,[7] which is described mathematically using homotopy theory. We shall have much more to say about this in the next chapter, but the main result is that, in many models, the solitons may be assigned an integer-valued "topological charge", which is conserved as the soliton evolves in time. Moreover, in a collision of two or more solitons the total charge is also conserved.

The topological charge has a natural physical interpretation. One can think of the solitons as subatomic particles, and of the topological charge as one of the conserved quantities of particle physics. The most successful model of this type is the famous model introduced by Skyrme almost 30 years ago.[8] At first it did not receive much attention, but interest has grown in recent years. It is a theory in (3+1) dimensions; the solitons are thought of as baryons, in particular protons and neutrons, and the topological charge is taken to be the conserved baryon number. These ideas lead to good qualitative agreement with many experimental results in baryon physics; a good review is contained in reference [9].

The Skyrme model is just one example of a σ-model appearing in physics. We now look briefly at the other important applications, which fall into two broad areas, namely particle physics and condensed matter physics. We begin with particle physics.

Although the Skyrme model was largely ignored when first proposed, σ-models *were* in fact popular in the early 1960s as models of the strong interactions between baryons, but in an alternative description, in which topology played no role. The baryons were no longer lumps of energy: instead they were represented within the framework of quantum theory by the nonlinear fields themselves. Their interactions were investigated using the techniques of current algebra. These models were popular for several years, although it was difficult to obtain any quantitative predictions from them.

In the late 1970s the σ-models reappeared as "toy models" in which to investigate the topological structure of Yang-Mills theories. This interest was sparked by the suggestion that topological effects might be connected with quark confinement. It was found that taking proper account of topology in the σ-models could indeed lead to confinement, which gave increased hope that the same thing might happen in QCD. In fact this has never been fully demonstrated, but there is now no doubt that topology plays a very important role in any model that is based on Yang-Mills fields.

To summarize, σ-models have appeared in particle physics firstly as a possible description of baryon physics and second as low-dimensional analogues of Yang-Mills fields. In both these applications they were treated mainly as quantum theories, in contrast with Skyrme's idea that they can be treated in a purely classical way, with the particle interpretation coming from the topological charge.

Turning now to condensed matter physics, one σ-model has proved to be of particular importance: this is the O(3) model, which is the main subject of chapters IV, V and VI. Classically it is the continuum limit of the Heisenberg ferromagnet; and recently it has been suggested that the quantum version of the model in (2+1) dimensions may provide mechanisms for high temperature superconductivity[10] and also for the quantum Hall effect.[11] At the centre of this latest speculation is the intriguing possibility of fractional statistics,[12] which are neither bosonic nor fermionic and are peculiar to quantum theories in (2+1) dimensions.

Many of these physical applications are discussed in more detail in chapter 10 of the recent book by Zakrzewski.[13] For the sake of completeness we remark that σ-models are also important in pure mathematics as examples of harmonic maps. Many of their properties have a natural description in the language of differential geometry.[14]

Finally, it should be stressed that this thesis deals exclusively with the *classical* theory of σ-models. Since most physical applications employ *quantum* theories, the most direct

relevance will not be in these areas, but in more mathematical areas, especially the theory of solitons. On the other hand, classical theories may be thought of as just the first approximation to the corresponding quantum theory; so it is possible that the results presented here might well turn out to have consequences for physics. In light of the current interest in condensed matter physics, this remark applies particularly to the sections dealing with the O(3) model.

The main body of the thesis is laid out as follows. Chapter II contains a brief review showing how the σ-models in (2+1) dimensions fit into the overall framework of nonlinear field theory. To begin with, the general σ-model action and equations of motion are presented, along with the coset description of homogeneous target manifolds. The most familiar examples are discussed in a little more detail. It is then shown how Derrick's theorem helps to identify the most likely candidates in the search for models with soliton solutions; and how, in the particular case of (2+1) dimensions, it picks out the σ-models. The rest of the chapter is concerned with the central role played by the topology of the target manifold. We develop the basic homotopy theory that is necessary to carry out a topological classification of nonlinear field theories in general. It is seen that, in (2+1) dimensions, the Kähler σ-models are of particular interest from the topological point of view. We go on to construct a topological charge for these models, using the techniques of differential geometry and cohomology theory. Finally, the Bogomolny bound is introduced, and used to derive the instanton (*i.e.* static soliton) solutions explicitly. Throughout the discussion, comparisons are made with the ϕ^4 and sine-Gordon equations in (1+1) dimensions, and also with the Skyrme model in (3+1).

Chapter III deals with a particular example of a modified σ-model that is "integrable", in the sense that procedures exist for writing down explicit multisoliton solutions in (2+1) dimensions. Although the model does not have any obvious application in physics, it is very interesting from the soliton theoretic point of view. For example, two lump-like solitons can pass through each other with no change in velocity and no lasting deformation. This type of behaviour is familiar from other integrable systems in (2+1) dimensions. But there are also wave-like solitons, about which very little is known at present. The work presented here constructs these waves, and investigates their interactions. In particular, it

is found that the interaction of a wave and a lump is quite different from any interaction occurring in other integrable models.

The next three chapters concentrate on lump-like solitons in the (unmodified) O(3) σ-model. A central question is that of soliton stability. Since there is no natural scale in the model, the solitons can take any size (although they always have the same general shape). There is the possibility that as a result of small perturbations they could either expand indefinitely, eventually covering the whole plane, or else shrink, so becoming infinitely tall spikes. It seems that to answer these questions, one must evolve the soliton configurations numerically on a lattice, since the O(3) model is nonintegrable. An important consideration in any numerical evolution is the choice of initial data. Specifically, to look at small perturbations, one wants to begin with a "discrete" instanton, *i.e.* a configuration that is static *on the lattice*. Chapter IV shows that by taking proper account of the topological aspects in the theory, one is led to a natural (but new) evolution scheme containing explicit discrete instantons. Using them as the basis for a study of soliton stability, chapter V reaches the conclusion that the O(3) lumps are unstable.

Moving on to the second important question, that of soliton interactions, chapter VI attempts to provide some answers by using an approximate technique in which the solitons are assumed to be slowly moving. The method was first suggested as a means of investigating the scattering of magnetic monopoles,[15] but it is equally applicable to the σ-models. The basic idea is that one constructs all *static* k-lump solutions (a problem that is discussed towards the end of chapter II). These solutions depend on a finite number (n_k, say) of real parameters. When the kinetic energy is small, a good approximation to the full field dynamics is obtained by evolving just these n_k parameters. Chapter VI restricts attention to two-lump scattering in the O(3) model. In the simplest head-on collisions, the lumps emerge at 90° to their original direction of motion; but there are also found to be several other, more exotic possibilities. Moreover, it found that soliton interactions also exhibit the instabilities revealed in chapter V. After scattering, the lumps either shrink to spikes or expand indefinitely, depending on their initial configuration.

Finally, chapter VII outlines work currently in progress, and also suggests some possible avenues for future research, looking in particular at modifications to the O(3) σ-model designed to stabilize the soliton against large changes in its size.

Chapter II
Sigma Models and their Topology

2.1 General Structure

The σ-models are a particular type of nonlinear scalar field theory, in which the scalar field takes values in an n-dimensional Riemannian manifold $\mathcal{M}$. Throughout this work it is assumed that they are defined on a $(d+1)$-dimensional Minkowski spacetime, with metric $\mathrm{diag}(-1,1,1,\ldots,1)$, although it is possible to consider them in more general geometries. The field equations are derived from the action

$$S = \frac{1}{4}\iint g_{ij}(\phi)\,\partial_\mu\phi^i\,\partial^\mu\phi^j\,d^dx\,dt\,, \tag{2.1}$$

where ϕ^i $(i = 1,2,\ldots,n)$ and g_{ij} are the coordinates and metric on $\mathcal{M}$ respectively. The Greek indices take values $0,1,2,\ldots,d$ and label the spacetime coordinates. Explicitly, the field equations derived from (2.1) are

$$\partial_\mu\partial^\mu\phi^i + \Gamma^i{}_{jk}\,\partial_\mu\phi^j\,\partial^\mu\phi^k = 0\,, \tag{2.2}$$

where $\Gamma^i{}_{jk}$ are the usual Christoffel symbols associated with the metric g.

The field equations are in general nonlinear, owing to the presence of the term quadratic in the field derivatives. It is the distinguishing feature of σ-models that the nonlinearities arise purely from the curvature of the target manifold $\mathcal{M}$. One does not need to add extra terms to the Lagrangian to introduce nonlinearities by hand; they are already an intrinsic part of the theory. As we shall see in the next section, it is precisely this type of model that can lead to soliton solutions in two spatial dimensions. Furthermore, all these models are Lorentz invariant, in the sense of possessing an $\mathrm{SO}(d,1)$ symmetry acting on spacetime. These are just two of the features that make such theories so interesting to physicists.

An alternative formulation of the field equations involves taking a free field theory, containing m independent fields (where $m > n$), and then using Lagrange multipliers to impose $(m-n)$ constraints, thereby restricting the fields to lie on $\mathcal{M}$. In essence, one is thinking of the n-dimensional manifold $\mathcal{M}$ as being embedded in $\mathbb{R}^m$. As an example of this approach, consider the O(3) σ-model, in which $\mathcal{M}$ is just a two-sphere. It is common to use three real fields $\boldsymbol{\phi} \equiv (\phi_1, \phi_2, \phi_3)$, with the constraint $\boldsymbol{\phi}.\boldsymbol{\phi} = 1$. Introducing a Lagrange multiplier λ, the relevant Lagrangian is

$$\mathcal{L} = \tfrac{1}{4}\,\partial_\mu \boldsymbol{\phi}.\partial^\mu \boldsymbol{\phi} - \lambda(\boldsymbol{\phi}.\boldsymbol{\phi} - 1)\,, \tag{2.3}$$

which leads to the field equations

$$\partial_\mu \partial^\mu \boldsymbol{\phi} + (\partial_\mu \boldsymbol{\phi}.\partial^\mu \boldsymbol{\phi})\boldsymbol{\phi} = \mathbf{0}\,. \tag{2.4}$$

The O(3) model is the central topic of chapters IV, V and VI; and the material in chapter III is based on a closely related model, namely the SU(2) chiral model, which may be formulated as follows. Suppose one adds an extra field ϕ_0 to the O(3) model and replaces the constraint $\boldsymbol{\phi}.\boldsymbol{\phi} = 1$ with

$$\phi_0^2 + \boldsymbol{\phi}.\boldsymbol{\phi} = 1\,. \tag{2.5}$$

The fields now take values on the three-sphere, which is topologically just the group manifold of SU(2). This equivalence is made explicit by writing the fields as a 2×2 unitary matrix:

$$J = \phi_0 \mathbb{1} + i\boldsymbol{\sigma}.\boldsymbol{\phi}\,, \tag{2.6}$$

where $\boldsymbol{\sigma}$ are the usual Pauli matrices. The constraint (2.5) ensures that J has unit determinant. The σ-model action may now be written in the form

$$\begin{aligned} S = \;& \frac{1}{4}\iint \mathrm{Tr}\,(\partial_\mu J\, \partial^\mu J^{-1})\, d^d x\, dt \\ = \;& -\frac{1}{4}\iint \mathrm{Tr}\,\big((J^{-1}\partial_\mu J)(J^{-1}\partial^\mu J)\big)\, d^d x\, dt \end{aligned} \tag{2.7}$$

and the corresponding equations of motion are

$$\partial_\mu (J^{-1}\partial^\mu J) = 0\,. \tag{2.8}$$

In fact, the original σ-models of the early 1960s were very similar to this, although they used slightly different notation and were treated as quantum theories, in which the

fields were interpreted as creation operators: ϕ was usually called π and was the creation operator for pions, ϕ_0 was usually called σ and created so-called σ-particles (hence the name that is now attached to virtually any nonlinear scalar field theory of this type). As a final remark, note that the O(3) model can be thought of as a special case of the SU(2) model, obtained by requiring $\phi_0 = 0$. Geometrically this corresponds to restricting the fields to lie on the equator of SU(2).

We now briefly consider the σ-models in a more abstract, geometrical way. This provides for a more systematic classification. From now on we shall assume that $\mathcal{M}$ is a homogeneous space, with a continuous group of symmetries G acting transitively on it. All the σ-models that are important in physics have this property, and it leads to a useful coset description for $\mathcal{M}$, in the following way.

First of all, pick a base point $y_0 \in \mathcal{M}$ and then construct the little group H of y_0 (sometimes called the isotropy subgroup), defined by

$$H \equiv \{h \in G \,:\, h(y_0) = y_0\}\,. \tag{2.9}$$

In other words, H is the set of elements of G that act trivially on y_0. Now suppose that g_1 and g_2 have the same action on y_0, *i.e.*

$$\begin{aligned} & g_1(y_0) = g_2(y_0) \\ \Rightarrow \quad & g_1^{-1} g_2(y_0) = y_0 \\ \Rightarrow \quad & g_1^{-1} g_2 \in H\,. \end{aligned}$$

Therefore g_1 and g_2 belong to the same left coset of G with respect to H. Clearly the converse also holds, namely two elements in the same left coset have the same action on y_0. Moreover, since G acts transitively, any point of $\mathcal{M}$ may be obtained as the action of some left coset. Hence we can make the identification

$$\mathcal{M} = G/H = \{gH \,:\, g \in G\}\,. \tag{2.10}$$

Note that since $\mathcal{M}$ is homogeneous, this construction is independent of the choice of y_0. Depending on the precise structure of $\mathcal{M}$, the model may be referred to in slightly different ways. In particular, if $\mathcal{M}$ is itself a Lie group, it is often called a "chiral model" (we have already met the SU(2) chiral model). This corresponds to H being trivial.

We now show explicitly how the coset description applies to three of the most important families of σ-models, namely the O(n) models, the $\mathbb{C}\mathrm{P}^n$ models[16,17] and the grassmannian models.[18]

O(n) Models

In the O(n) model the fields take values on the $(n-1)$-sphere, for which the symmetry group is just the connected group of rotations in n dimensions, namely SO(n). Having picked a base point y_0, the rotations that leave it invariant are those about the direction of y_0 itself; so H is simply SO($n-1$). Hence we have

$$\begin{aligned} &\mathcal{M} = \mathrm{SO}(n)/\mathrm{SO}(n-1) = S^{n-1}\,, \\ &\dim \mathcal{M} = n-1\,. \end{aligned} \tag{2.11}$$

The O(n) model can be formulated in terms of $n-1$ real fields, but as already pointed out, it is more common to use n fields, together with a single constraint that forces them to lie on S^{n-1}.

CP(n) Models

In these models the fields take values in the complex projective space $\mathbb{C}\mathrm{P}^n$. In terms of the coset description we have

$$\begin{aligned} &\mathcal{M} = \mathrm{SU}(n+1)/\big(\mathrm{SU}(n)\times \mathrm{U}(1)\big) = \mathbb{C}\mathrm{P}^n\,, \\ &\dim \mathcal{M} = 2n\,. \end{aligned} \tag{2.12}$$

One can use either $2n$ real fields, or alternatively n complex fields. It is interesting to note that $\mathbb{C}\mathrm{P}^1$ is isometric to S^2, so that the $\mathbb{C}\mathrm{P}^1$ model is nothing more than the O(3) model in disguise. This equivalence may be made explicit by taking

$$u = \frac{\phi_1 + i\phi_2}{1-\phi_3}\,, \tag{2.13}$$

where u is the (complex) $\mathbb{C}\mathrm{P}^1$ field and (ϕ_1, ϕ_2, ϕ_3) are the usual O(3) fields. However, there is no similar equivalence for higher values of n. One can take the view that the O(n) and $\mathbb{C}\mathrm{P}^n$ models are two distinct extensions of the O(3) model to higher dimensional target spaces.

Grassmannian Models

In these models the fields take values in the complex grassmannian manifolds $G_{m,n}$. In terms of the coset description we have

$$\begin{aligned} &\mathcal{M} = \mathrm{SU}(m+n)/\big(\mathrm{SU}(m) \times \mathrm{SU}(n) \times \mathrm{U}(1)\big) = G_{m,n}\,, \\ &\dim \mathcal{M} = 2mn\,. \end{aligned} \tag{2.14}$$

One can use either $2mn$ real fields, or alternatively mn complex fields. Note that $G_{1,n} = \mathbb{C}\mathrm{P}^n$, and so the grassmannian models are a natural generalization of the $\mathbb{C}\mathrm{P}^n$ models, containing them as a subset.

Finally in this section, we make a few remarks about Kähler σ-models.[19] It turns out that if $\mathcal{M}$ admits a Kähler metric then the equations of motion are considerably simplified. Both the grassmannian and $\mathbb{C}\mathrm{P}^n$ models have this property. To see how things work in general, suppose that $\mathcal{M}$ is an n-dimensional complex manifold (*i.e.* $2n$ real dimensions), parametrized by coordinates u^α together with their complex conjugates $\bar{u}^\alpha$, with α taking values $1, 2, \ldots, n$. The σ-model action may be written

$$S = \frac{1}{4} \iint g_{\alpha\bar{\beta}}(u)\, \partial_\mu u^\alpha\, \partial^\mu \bar{u}^\beta\, d^d x\, dt\,, \tag{2.15}$$

where the barred subscript on the metric g labels conjugate coordinates. The model is said to be Kähler if the two-form

$$\omega \equiv g_{\alpha\bar{\beta}}\, \mathrm{d}u^\alpha \wedge \mathrm{d}\bar{u}^\beta \tag{2.16}$$

is closed, *i.e.* $\mathrm{d}\omega = 0$. This condition implies that all the mixed Christoffel symbols vanish, *i.e.*

$$\Gamma^\alpha{}_{\beta\bar{\gamma}} = \Gamma^\alpha{}_{\bar{\beta}\bar{\gamma}} = \Gamma^{\bar{\alpha}}{}_{\beta\bar{\gamma}} = \Gamma^{\bar{\alpha}}{}_{\beta\gamma} = 0\,. \tag{2.17}$$

The equations of motions (2.2) are greatly simplified as a result, and may now be written in the form

$$\partial_\mu \partial^\mu u^\alpha + g^{\alpha\bar{\delta}} g_{\beta\bar{\delta},\gamma}\, \partial_\mu u^\beta\, \partial^\mu u^\gamma = 0\,, \tag{2.18}$$

where $g_{\beta\bar{\delta},\gamma}$ denotes the partial derivative of $g_{\beta\bar{\delta}}$ with respect to u^γ. As an example consider the $\mathbb{C}\mathrm{P}^1$ model. There is just one complex field u (so the indices α and β may

be dropped); u is related to the O(3) fields by (2.13). Choosing $g = 4(1+u\bar{u})^{-2}$ in (2.15) yields the conventional action of the O(3) model, and then from (2.18) the equation of motion for u is

$$\partial_\mu \partial^\mu u - \frac{2\bar{u}\,\partial_\mu u\,\partial^\mu u}{1+u\bar{u}} = 0\,. \tag{2.19}$$

2.2 Derrick's Theorem

For both mathematicians and physicists, it would be very useful to have a method of determining, solely from the form of the Lagrangian, whether a given theory possesses soliton solutions. At the moment such a goal seems a little out of reach; but there is an important result, due to Derrick,[20] which for a great many models can say when solitons will definitely *not* occur. It is instructive to go through the proof of the theorem, first because the method of argument is applicable to a wide range of other problems, and second because it gives an opportunity to introduce some new field theories, which, although not σ-models in the strict sense, nevertheless have many features in common with them.

Consider the (Lorentz invariant) nonlinear field theories, living in (d+1)-dimensional Minkowski spacetime, that are defined by the Lagrangian

$$\mathcal{L} = \tfrac{1}{4} g_{ij}\,\partial_\mu \phi^i\,\partial^\mu \phi^j + V[\boldsymbol{\phi}]\,. \tag{2.20}$$

These models may be thought of as an extension of the σ-models, obtained by allowing an explicit potential term $V[\boldsymbol{\phi}]$. To ensure that the energy density is positive definite, $V[\boldsymbol{\phi}]$ is always assumed non-negative. The energy E may be split into two pieces:

$$E[\boldsymbol{\phi}] = E_1[\boldsymbol{\phi}] + E_2[\boldsymbol{\phi}]\,, \tag{2.21}$$

where

$$\begin{aligned} E_1[\boldsymbol{\phi}] &= \frac{1}{4}\int g_{ij}\partial_\mu \phi^i \partial^\mu \phi^j\, d^d x\,, \\ E_2[\boldsymbol{\phi}] &= \int V[\boldsymbol{\phi}]\, d^d x\,. \end{aligned} \tag{2.22}$$

From the point of view of soliton theory, one is particularly interested in the possible existence of static lump solutions, localized in the sense that E is finite. Moving lumps

are then obtained from the static ones simply by Lorentz boosting. Suppose then that $\boldsymbol{\phi}(\mathbf{r})$ is a static lump solution where $\mathbf{r}$ is just shorthand for $(x^1, x^2, \ldots, x^d)$.

Derrick showed that the difference in the scaling behaviours of E_1 and E_2 places certain restrictions on the theories that have such solutions. Consider the one-parameter family of configurations

$$\boldsymbol{\phi}_\lambda(\mathbf{r}) \equiv \boldsymbol{\phi}(\lambda \mathbf{r}) . \tag{2.23}$$

It is easy to see that E_1 and E_2 scale in such a way that the energy of these configurations is

$$E_\lambda \equiv E[\boldsymbol{\phi}_\lambda] = \lambda^{2-d} E_1[\boldsymbol{\phi}] + \lambda^{-d} E_2[\boldsymbol{\phi}] . \tag{2.24}$$

The crucial point is to note that $\boldsymbol{\phi}(\mathbf{r})$ is a static solution only if it is an extremum of E, and in particular only if it makes E_λ stationary with respect to variations in λ. In other words $dE_\lambda/d\lambda$ must vanish when evaluated at $\lambda = 1$. In general we have

$$\frac{dE_\lambda}{d\lambda} = (2-d)\lambda^{1-d} E_1[\boldsymbol{\phi}] - d\lambda^{-d-1} E_2[\boldsymbol{\phi}] \tag{2.25}$$

and so the following cases can occur, depending on the number of spatial dimensions.

$$d = 1 : \quad \left.\frac{dE_\lambda}{d\lambda}\right|_{\lambda=1} = 0 \quad \Rightarrow \quad E_1[\boldsymbol{\phi}] - E_2[\boldsymbol{\phi}] = 0 .$$

$$d = 2 : \quad \left.\frac{dE_\lambda}{d\lambda}\right|_{\lambda=1} = 0 \quad \Rightarrow \quad E_2[\boldsymbol{\phi}] = 0 .$$

$$d > 2 : \quad \left.\frac{dE_\lambda}{d\lambda}\right|_{\lambda=1} \quad \text{is always negative.}$$

The following conclusions can be drawn. First, in one spatial dimension the potential term *must* be present if there are to be soliton solutions. But in two dimensions the exact opposite is true, *i.e.* if the potential term is included then solitons certainly do *not* exist. As far as the σ-models are concerned, this means that in (1+1) dimensions they are not very interesting, but in (2+1) they are precisely the models that one should consider in order to find examples of relativistic solitons. Finally, in 3 (or more) spatial dimensions things look a little bleak: there is no soliton in any theory of the form (2.20). However, three-dimensional solitons do occur in slightly more general models, and we shall return to this point shortly.

The previous paragraph summarizes the main content of Derrick's theorem, but it is also instructive to look at the second derivative of E_λ, since this gives information about soliton stability. It is easy to check that

$$d = 1 : \quad \left.\frac{d^2 E_\lambda}{d\lambda^2}\right|_{\lambda=1} = 2E_2[\phi] > 0\,,$$

$$d = 2 : \quad \left.\frac{d^2 E_\lambda}{d\lambda^2}\right|_{\lambda=1} = 6E_2[\phi] = 0\,.$$

Therefore, in one spatial dimension, $\lambda = 1$ corresponds to a minimum of E, *i.e.* the soliton has a preferred scale and resists attempts to change its size. However, in two dimensions the situation is somewhat different, since the theory now contains a zero mode. In this case there is no preferred scale and it is possible that the soliton could either shrink or expand under small perturbations. This feature is a consequence of the conformal invariance of the σ-models in two spatial dimensions, and is studied in some depth in chapters IV and V.

We now briefly mention two models, both living in (1+1) dimensions, that have soliton solutions (often called kinks when $d = 1$). They are both of the form (2.20) (and so under the terms of Derrick's theorem have nonzero potential terms) and each contains just a single real field ϕ.

Phi-fourth Equation

$$\begin{aligned}
&\text{Lagrangian} && \mathcal{L} = \tfrac{1}{4}\partial_\mu\phi\,\partial^\mu\phi + \tfrac{1}{8}(\phi^2-1)^2\,,\\
&\text{field equation} && \phi_{xx} - \phi_{tt} + \phi(1-\phi^2) = 0\,,\\
&\text{kink solutions} && \phi = \pm\tanh\Big(\frac{1}{\sqrt{2}}(x-x_0)\Big)\,.
\end{aligned} \tag{2.26}$$

Sine-Gordon Equation

Lagrangian	$\mathcal{L} = \frac{1}{4}\partial_\mu\phi\,\partial^\mu\phi + \frac{1}{2}(1-\cos\phi)\,,$
field equation	$\phi_{xx} - \phi_{tt} - \sin\phi = 0\,,$
kink solutions	$\phi = \pm 4\arctan\big(\exp(x-x_0)\big)\,.$

(2.27)

To end this section we discuss ways of side-stepping Derrick's theorem in three spatial dimensions. The first possibility is to change slightly what one means by a static solution. For some models it is possible write down field configurations that are themselves dependent on time, but which have a time-independent energy density. Just one example of such objects are the so-called Q-balls, which were originally introduced by Coleman.[21] Another option is to move outside the realm of scalar field theories and to allow the inclusion of gauge fields. Of particular interest are theories in which a Yang-Mills gauge field is coupled to a scalar Higgs field — these are the theories in which magnetic monopoles occur.

The third possibility is to stay with scalar fields, but to include an extra term in the Lagrangian. This is the idea behind the Skyrme model,[8] which may be constructed as a modification of the SU(2) chiral model in the following way. Recall that the fields may be written as an SU(2)-valued matrix J. The term added to the Lagrangian takes the form

$$\mathcal{L}_{\text{Skyrme}} = -\frac{1}{32e^2}\,\text{Tr}\left([J^{-1}\partial_\mu J\,,\,J^{-1}\partial_\nu J]\,[J^{-1}\partial^\mu J\,,\,J^{-1}\partial^\nu J]\right), \tag{2.28}$$

where e^2 has the physical interpretation of a coupling constant. The extra term gives a (positive) contribution E_{Skyrme} to the potential energy. The scaling behaviour of the total potential energy, *i.e.* the analogue of (2.24), is now

$$E_\lambda = \lambda^{2-d}\,E_1[J] + \lambda^{-d}\,E_2[J] + \lambda^{4-d}\,E_{\text{Skyrme}}[J]\,, \tag{2.29}$$

and differentiating this in the case $d = 3$ yields

$$\frac{dE_\lambda}{d\lambda} = -\lambda^{-2}\,E_1[J] - 3\lambda^{-4}\,E_2[J] + E_{\text{Skyrme}}[J]\,. \tag{2.30}$$

It is conventional to set the potential term to zero, *i.e.* $E_2 = 0$, in which case there is a

stationary point of E_λ at

$$\lambda^2 = \frac{E_1}{E_{\text{Skyrme}}}. \tag{2.31}$$

In contrast with the pure σ-model there is the possibility of a soliton solution. In fact such a solution *does* exist and, not unnaturally, it has become known as the skyrmion. Its main importance is in playing the role of baryons in the Skyrme model of low-energy hadronic physics. Unfortunately the skyrmion configuration cannot be found analytically, and so its properties must be investigated numerically. Nevertheless, it is without doubt the most important example of a soliton having a direct interpretation in particle physics.

2.3 Homotopy Theory

So far we have looked at the general structure of the field equations for σ-models, and in doing so have been concerned mainly with the local properties of the target manifold $\mathcal{M}$. In the remainder of the chapter we shall emphasize the global topology of $\mathcal{M}$, which is very important in the classification of solitons and leads eventually to their explicit construction. To do this we shall first need to gather together some basic results in homotopy theory.[7]

Roughly speaking, one says that two continuous maps are homotopic if they can be continuously deformed one into the other. More precisely, if X and Y are two manifolds, and f and g are two continuous maps from X to Y, then f and g are homotopic if there exists a continuous map

$$h : X \times [0,1] \to Y$$

such that for all $x \in X$

$$h(x,0) = f(x),$$
$$h(x,1) = g(x).$$

It is easy to verify that the notion of homotopy defines an equivalence relation on the set of all continuous maps from X to Y. Hence the maps are partitioned into homotopy classes, with any two maps belonging to the same class if and only if they are homotopic. It is conventional to denote the homotopy class of f by $[f]$.

It is possible to define a binary operation between homotopy classes such that they take on a group structure. In general this involves a certain amount of technical detail,

but it is easy to illustrate the general procedure in a simple example. Suppose that X is just a circle S^1, parametrized by θ in the range $[0, 2\pi]$. Then continuous maps from X to Y may be thought of as closed loops in Y. Consider only those maps that take a particular point of X to a particular point in Y; specifically take

$$f : [0, 2\pi] \to Y \quad \text{with} \quad f(0) = f(2\pi) = y_0 ,$$
$$g : [0, 2\pi] \to Y \quad \text{with} \quad g(0) = g(2\pi) = y_0 .$$

Now define a third map $h : X \to Y$ by

$$h(\theta) = \begin{cases} f(2\theta) & \text{if } 0 \leq \theta \leq \pi . \\ g(2(\theta - \pi)) & \text{if } \pi \leq \theta \leq 2\pi . \end{cases}$$

Clearly h is continuous and also satisfies $h(0) = h(2\pi) = y_0$. It is conventional to write $h = f + g$. The crucial point is to realise that the homotopy class of h depends only on the homotopy classes of f and g, and so this binary operation between maps induces a binary operation between homotopy classes. In an obvious notation we write

$$[h] = [f + g] = [f] + [g] .$$

It is not difficult to convince oneself that, under this operation, the set of all homotopy classes now forms a group, conventionally denoted by $\pi_1(Y, y_0)$ (the subscript 1 refers to the fact that X was taken to be a one-sphere). If Y has a continuous symmetry group G, acting transitively on it, then the homotopy group is independent of the choice of y_0 and one writes simply $\pi_1(Y)$. The whole construction described above can be generalized to the case where X is an n-sphere (see, for example, reference [7]), and the homotopy group is then denoted $\pi_n(Y)$.

The first two homotopy groups have a simple geometrical interpretation. Firstly, since S^0 consists of just two points, $\pi_0(Y)$ counts the number of disjoint pieces of Y. In particular, $\pi_0(Y) = 0$ if and only if Y is connected. Secondly, as suggested above, $\pi_1(Y)$ classifies the set of closed loops in Y. In particular, $\pi_1(Y) = 0$ if and only if Y is *simply* connected. In general the higher homotopy groups do not have such a natural interpretation, but they still play important roles in the context of field theory.

We now make the connection between homotopy theory and σ-models. The key observation is that (in two or more spatial dimensions) the requirement of finite potential energy means that the fields must tend to the *same* value at spatial infinity, regardless of direction. Hence the spatial degrees of freedom may be compactified from $\mathbb{R}^d$ to S^d. Each possible field configuration may be thought of as a map

$$\phi : S^d \to \mathcal{M} .$$

It is now clear that the configurations are classified by the homotopy group $\pi_d(\mathcal{M})$. As the fields evolve in time they do so continuously, and so ϕ always remains within the same homotopy class (sometimes referred to as a topological sector). If all configurations are homotopic, *i.e.* if $\pi_d(\mathcal{M}) = 0$, then the model is said to be topologically trivial. Topologically nontrivial theories enjoy a topological stability, in the sense that each configuration cannot evolve out of its original topological sector.

From Derrick's theorem it is known that, as far as the σ-models are concerned, the most interesting behaviour will occur in two spatial dimensions, and so one should look at $\pi_2(\mathcal{M})$. Rather than go into the technical details of calculating homotopy groups, we shall just state the results for the σ-models that we have encountered so far. To begin with

$$\begin{aligned} \pi_n(S^n) &= \mathbb{Z} , \\ \pi_n(S^m) &= 0 \quad (m > n) . \end{aligned} \tag{2.32}$$

The first of these is the formal statement that a map from S^n to S^n is characterized by a winding number; the second is a generalization of the intuitive idea that a closed loop can "slip off" a two-sphere. The consequence for σ-models is that, in d spatial dimensions, the O(d+1) model is topologically nontrivial, but that all higher ones are trivial. In particular, the (2+1)-dimensional O(3) model is nontrivial.

The next result is especially concise: all the chiral models in two spatial dimensions are trivial, because for any Lie group G we have

$$\pi_2(G) = 0 . \tag{2.33}$$

It remains to look at the grassmannian models. Here one needs to use two basic

results in homotopy theory:

$$\pi_n(X \times Y) = \pi_n(X) \oplus \pi_n(Y)\,, \tag{2.34}$$

$$\pi_2(X/Y) = \pi_1(Y) \quad \text{provided} \quad \pi_0(X) = \pi_1(X) = 0\,. \tag{2.35}$$

Since SU(n) is both connected and simply connected, (2.35) shows that the topological content of the grassmannian models is given by the first homotopy group of the denominator H in the coset description of the target manifold. Noting that $\pi_1(\mathrm{U}(1)) = \mathbb{Z}$ (since U(1) is topologically just S^1) and using (2.34) to calculate the homotopy group of a direct product, one finds

$$\pi_2(G_{m,n}) = \pi_1\big(\mathrm{SU}(m) \times \mathrm{SU}(n) \times \mathrm{U}(1)\big) = \mathbb{Z}\,. \tag{2.36}$$

In fact, for any compact Kähler symmetric space $\mathcal{M}$, $\pi_2(\mathcal{M}) = \mathbb{Z}$, *i.e.* all such models are topologically nontrivial. This is another reason why the Kähler σ-models are of particular interest.

In each of the above examples of nontrivial topology, the homotopy group is isomorphic to $\mathbb{Z}$, the additive group of integers. It would be nice to make this isomorphism a little more concrete by labelling each homotopy class with an integer-valued "topological charge". To be of any real use, such a charge must be constructed so that it is invariant as the fields evolve in time. This problem is addressed in the next section. In some (exotic) cases, one would require two or more separate charges. For example, consider the flag space[22]

$$F_n = \mathrm{SU}(n+1)\,/\,\mathrm{U}(1) \times \cdots \times \mathrm{U}(1)\,, \tag{2.37}$$

where there are n factors of U(1) in the denominator. Using the results (2.34) and (2.35) as before, it is easy to see that

$$\pi_2(F_n) = \mathbb{Z} \oplus \mathbb{Z} \oplus \cdots \oplus \mathbb{Z} \qquad (n \text{ times})\,. \tag{2.38}$$

Therefore a σ-model with fields taking values in F_n would need n distinct topological charges to label its homotopy classes.

So far we have said nothing about the construction of solutions to any particular σ-model; but we have seen that the combination of Derrick's theorem and a little homotopy theory yields a great deal of information from very general assumptions. To summarize, in two spatial dimensions the σ-models are precisely the scalar field theories that are capable of supporting soliton solutions. Moreover, in the case of the Kähler models it is expected that the solitons would be topologically stable (*i.e.* the soliton configurations are not homotopic to the vacuum). The solitons of the Kähler models will be constructed in section 2.5. However, it should be stressed that nontrivial topology does not guarantee the existence of solitons; as a counterexample consider the SU(2) chiral model (equivalently the O(4) σ-model) in (3+1) dimensions. Conversely, solitons *can* occur in models that are topologically trivial (for example the SU(2) model in (2+1) dimensions).

2.4 The Topological Charge

In this section we construct the topological charge for the Kähler models in (2+1) dimensions. The most elegant approach uses differential geometry to take account of the Kähler property of the target manifold, and to do this requires a little of the machinery of differential forms.[23]

Recall that a p-form α on a manifold $\mathcal{M}$ is closed if $\mathrm{d}\alpha = 0$ and is exact if $\alpha = \mathrm{d}\beta$ for some $(p-1)$-form β. All exact forms are closed, since $\mathrm{d}^2 = 0$; but all closed forms are exact only if $\mathcal{M}$ is contractible (this result is essentially the Poincaré Lemma). In general the closed p-forms on $\mathcal{M}$ are partitioned into so-called cohomology classes, with a pair of forms being in the same class if and only if they differ by an exact form. It is conventional to denote the set of closed p-forms by $Z^p(\mathcal{M})$ and the exact p-forms by $B^p(\mathcal{M})$. Then the set of cohomology classes may be written as a quotient space:

$$H^p(\mathcal{M}) = Z^p(\mathcal{M}) \,/\, B^p(\mathcal{M}) \,. \tag{2.39}$$

Because the sum of two closed forms is closed and the sum of two exact forms is exact, addition induces a group operation on $H^p(\mathcal{M})$, which is then known as the pth cohomology group of $\mathcal{M}$.

Although cohomology is defined *locally* in terms of the exterior derivative, it actually conceals a great deal of information about the *global* topology of $\mathcal{M}$. This hidden content

is made explicit by the Hurewicz theorem. In its simplest version, this says that if $\mathcal{M}$ is both connected and simply connected with the lowest nonzero homotopy group $\pi_n(\mathcal{M})$, then $H^n(\mathcal{M}) = \pi_n(\mathcal{M})$ and all lower cohomology groups are zero. In view of this result it is not so surprising that cohomology proves to be such a useful tool in the construction of a topological charge.

The necessary groundwork is completed with the notion of a pullback mapping. Suppose that $T_p(X)$ and $T_p^*(X)$ are the tangent and cotangent spaces respectively at some point p of a (real) manifold X. Consider a vector V in $T_p(X)$ and a one-form α in $T_p^*(X)$, given explicitly by

$$V = b^i \frac{\partial}{\partial x^i}\,, \quad \alpha = a_i \, \mathrm{d}x^i\,.$$

There is an inner product, defined so that

$$\langle \alpha, V \rangle = a_i b^i\,,$$

and if f is a real-valued function on X, then V acts on f according to

$$Vf \equiv \langle \mathrm{d}f, V \rangle = b^i \frac{\partial f}{\partial x^i}\,.$$

Now introduce a second manifold Y. Suppose that g is a real-valued function on Y, that ϕ is a mapping from X to Y, and that ω is a one-form living in $T^*_{\phi(p)}(Y)$. Then ϕ induces maps

$$\begin{aligned} &\phi_* : T_p(X) \to T_{\phi(p)}(Y) \quad \text{defined by} \quad (\phi_* V)g = Vg\big(\phi(p)\big)\,, \\ &\phi^* : T^*_{\phi(p)}(Y) \to T_p^*(X) \quad \text{defined by} \quad \langle \phi^*\omega, V \rangle = \langle \omega, \phi_* V \rangle\,. \end{aligned} \tag{2.40}$$

It is this last map, ϕ^*, that is the pullback mapping, taking one-forms on Y into one-forms on X. These ideas easily generalize to pullback mappings between r-forms.

The most important property of ϕ^* is that it takes closed forms to closed forms and exact forms to exact forms, and so may be considered as a map between cohomology classes. In the (2+1)-dimensional σ-models, one views the field configurations as mappings

$\phi : S^2 \to \mathcal{M}$. So each configuration defines a pullback mapping

$$\phi^* : H^2(\mathcal{M}) \to H^2(S^2). \tag{2.41}$$

It is at this point that the Kähler form

$$\omega = g_{\alpha\bar{\beta}}\, \mathrm{d}u^\alpha \wedge \mathrm{d}\bar{u}^\beta \tag{2.42}$$

comes into play. Since ω is closed it lies in some cohomology class $[\omega]$ (this notation should not lead to confusion with homotopy classes). The topological charge may now be defined by

$$Q(\phi) = c^{-1} \int_{S^2} \phi^*([\omega]), \tag{2.43}$$

where c is a normalization constant chosen so that Q takes on integer values. It is easy to see that Q is well defined. Because S^2 has no boundary, the integral of any exact form over S^2 vanishes (by Stokes' theorem) and so it makes sense to talk about the integral of a cohomology class over S^2.

It only remains to prove that Q is invariant as the fields evolve in time. There is a simple theorem in differential geometry,[23] stating that if ϕ_1 and ϕ_2 are homotopic then the pullbacks ϕ_1^* and ϕ_2^* are *equal* (when considered as maps between cohomology classes). We have already noted that, as the fields evolve, ϕ remains in the same homotopy class, *i.e.* ϕ^* is unchanged. Hence Q is invariant.

It should be stressed that Q is fundamentally different from the conserved quantities that arise as a result of continuous symmetries of the Lagrangian (for example a conserved energy arising as the result of time translation invariance). The existence of Q is due solely to the compactification of the spatial dimensions into a two-sphere, *i.e.* it is a consequence of the boundary conditions imposed to ensure finite energy.

It is possible to recast Q in terms of a particular set of manifold coordinates. In the (2+1)-dimensional Kähler models one finds

$$Q = ic^{-1} \int g_{\alpha\bar{\beta}}\, \epsilon_{ij}\, \partial_i u^\alpha\, \partial_j \bar{u}^\beta\, d^2x, \tag{2.44}$$

where ϵ_{ij} is the antisymmetric symbol on two indices such that $\epsilon_{12} = 1$. As an illustration take the $\mathbb{C}\mathrm{P}^1$ model, with $g = 4(1 + u\bar{u})^{-2}$ as before. It turns out that one must have

$c = 8\pi$ to obtain an integer charge:

$$Q = \frac{i}{2\pi} \int \frac{\partial_x u \partial_y \bar{u} - \partial_y u \partial_x \bar{u}}{(1 + u\bar{u})^2} \, d^2x \, . \tag{2.45}$$

If one now considers the equivalence with the O(3) model, and relates u to $\boldsymbol{\phi}$ according to (2.13), then Q may be written in the form

$$Q = -\frac{1}{8\pi} \int \epsilon_{ij} \boldsymbol{\phi} \, . \, (\partial_i \boldsymbol{\phi} \times \partial_j \boldsymbol{\phi}) \, \mathrm{d}^2x \, . \tag{2.46}$$

We now turn briefly to the topological content of theories living in (1+1) dimensions.[24] In this case the spatial boundary consists of just two points, namely $x = \pm\infty$. Finite energy requirements dictate that at these points the field should be at a minimum of the potential $V[\phi]$, but not necessarily at the same one (*i.e.* there is no justification for compactifying space as in higher dimensions). Assume that V has a discrete set of minima (possibly infinite). Then because the fields evolve continuously, ϕ must remain constant at infinity, *i.e.* it cannot jump from one minimum to another without violating finite energy. Let us define

$$\begin{aligned} \phi_+ &\equiv \phi(+\infty, t) \, , \\ \phi_- &\equiv \phi(-\infty, t) \, . \end{aligned} \tag{2.47}$$

Again the finite-energy configurations are partitioned into topological sectors, in the sense that a configuration with particular values of ϕ_+ and ϕ_- cannot evolve into one with different values. In the simplest case $V[\phi]$ has only one minimum and there is just one topological sector, but in general the situation is more interesting.

For example, in the ϕ^4 equation (2.26) there are two minima of $V[\phi]$, at $\phi = \pm 1$. Hence there are four topological sectors. It is possible to label them with an integer-valued topological index, given by

$$Q = \tfrac{1}{2}(\phi_+ - \phi_-) \, . \tag{2.48}$$

The solutions written down in (2.26) have $Q = \pm 1$; these are conventionally called the kink and antikink respectively. The other two sectors both have $Q = 0$.

The sine-Gordon equation (2.27) has yet a little more structure, since now $V[\phi]$ has a infinite number of minima, at $\phi = 2n\pi$. In this case the topological index is given by

$$Q = \tfrac{1}{2\pi}(\phi_+ - \phi_-)\,. \tag{2.49}$$

There are an infinite number of topological sectors, with Q able to take any integer value. For example, the kink and antikink solutions given in (2.27) have $Q = \pm 1$. However, static solutions do not exist in all sectors, because of a simple argument[24] showing that any nontrivial static solution must connect adjacent minima. For the ϕ^4 equation this result is irrelevant, since there are only two minima anyway; but for potentials with more minima, such as sine-Gordon, it means that there are no static solutions with charge two or more. In other words, the kink and antikink of (2.27) are the only static solutions.

2.5 Bogomolny Bounds

In this section it is shown how to construct static solutions to most of the topologically nontrivial field theories that we have discussed so far, using a general technique first suggested by the Russian physicist Bogomolny.[25] Along the way one obtains a lower bound on the potential energy in each topological sector. To begin with consider the scalar field theories in (1+1) dimensions given by the Lagrangian

$$\mathcal{L} = \tfrac{1}{4}\partial_\mu\phi\,\partial^\mu\phi + V[\phi]\,. \tag{2.50}$$

The potential energy in these models is simply

$$\begin{aligned} E &= \int_{-\infty}^{+\infty} \left(\tfrac{1}{4}\phi_x^2 + V[\phi]\right) dx \\ &= \int_{-\infty}^{+\infty} \left(\left(\tfrac{1}{2}\phi_x - \sqrt{V[\phi]}\right)^2 + \phi_x\sqrt{V[\phi]}\right) dx \\ &= \int_{-\infty}^{+\infty} \left(\tfrac{1}{2}\phi_x - \sqrt{V[\phi]}\right)^2 dx + \int_{\phi_-}^{\phi_+} \sqrt{V[\phi]}\, d\phi\,. \end{aligned}$$

Therefore

$$E \geq \int_{\phi_-}^{\phi_+} \sqrt{V[\phi]}\, d\phi\,, \quad \text{with equality if and only if} \quad \tfrac{1}{2}\phi_x = \sqrt{V[\phi]}\,. \tag{2.51}$$

This lower bound on E is known as the Bogomolny bound, and the condition for equality is the Bogomolny equation. For example in the $Q = 1$ sector of the ϕ^4 equation we find

$$E \geq \frac{1}{2\sqrt{2}} \int_{-1}^{+1} (1 - \phi^2)\, d\phi = \frac{\sqrt{2}}{3}\,,$$

with equality if and only if

$$\phi_x - \frac{1}{\sqrt{2}}(1 - \phi^2) = 0\,. \tag{2.52}$$

Similarly, for the sine-Gordon equation one has

$$E \geq \frac{1}{\sqrt{2}} \int_{0}^{2\pi} \sqrt{1 - \cos\phi}\, d\phi = 4\,,$$

with equality if and only if

$$\phi_x - \sqrt{2(1 - \cos\phi)} = 0\,. \tag{2.53}$$

Solving (2.52) and (2.53) simply yields the kink (and antikink) solutions (2.26) and (2.27) respectively.

At first sight the Bogomolny technique seems to involve nothing more than completing the square, and this is indeed the case; but it is no less powerful for that. The first thing to note is that the static solutions are obtained by solving a *first-order* equation, wheras the original field equation is always *second order*. In addition, their explicit form is not needed in order to find their energy: one need simply know that a solution to the Bogomolny equation exists.

These ideas may be extended to the Kähler σ-models in (2+1) dimensions.[19] In this case it is better to start from the identity

$$\int g_{\alpha\bar{\beta}} \left(\partial_i u^\alpha \pm i\epsilon_{ij}\partial_j u^\alpha\right)\left(\partial_i \bar{u}^\beta \mp i\epsilon_{ik}\partial_k \bar{u}^\beta\right) d^2x \geq 0\,. \tag{2.54}$$

Using the relations

$$\begin{aligned} E &= \frac{1}{4}\int g_{\alpha\bar{\beta}}\,\partial_i u^\alpha\,\partial_i \bar{u}^\beta\, d^2x\,, \\ cQ &= i\int g_{\alpha\bar{\beta}}\,\epsilon_{ij}\,\partial_i u^\alpha\,\partial_j \bar{u}^\beta\, d^2x\,, \end{aligned} \tag{2.55}$$

it is easy to recast (2.54) to give

$$E \geq \frac{1}{4}|cQ|\,, \tag{2.56}$$

with equality if and only if

$$\partial_i u^\alpha \pm i\epsilon_{ij}\partial_j u^\alpha = 0\,. \tag{2.57}$$

The Bogomolny equation (2.57) is sometimes referred to as the self-duality equation. It can be simplified dramatically by moving to a complex coordinate $z = x + iy$. If ∂ and $\bar{\partial}$ denote differentiation with respect to z and $\bar{z}$ respectively, then the upper sign in (2.57) leads to

$$\bar{\partial} u^\alpha = 0\,, \tag{2.58}$$

while the lower sign gives

$$\partial u^\alpha = 0\,. \tag{2.59}$$

In the first case u may be any analytic function of z; these solutions are called instantons and have a positive topological charge. In the second case u is an anti-analytic function of z and gives rise to anti-instantons, which have a negative topological charge.

The mathematical beauty of this formalism is undeniable. To see how things work in a particular example, consider once more the $\mathbb{C}\mathrm{P}^1$ model. Recall that $c = 8\pi$ and so the Bogomolny bound becomes

$$E \geq 2\pi|Q|\,. \tag{2.60}$$

The condition of finite energy restricts u to rational functions of z (instantons) or $\bar{z}$ (anti-instantons). The degree of u is equal to the absolute value of the topological charge. If

$|Q| = k$ then a solution of charge Q is a rational function of degree k, which has $4k+2$ real parameters. However, there is an invariance under global rotations of the target manifold, and this removes 3 parameters. Therefore a k-instanton depends on only $4k-1$ parameters. For example, when $k=1$ one may take

$$u = \frac{\lambda}{z-\mu}, \tag{2.61}$$

where λ is real but μ may be complex. The physical interpretation is that one has an instanton of size λ, centred around a point μ in the complex plane. For $k=2$ the instantons are given by

$$u = \frac{2\beta z + \gamma}{z^2 + \delta z + \epsilon}, \tag{2.62}$$

with β real and the other parameters complex; and likewise for higher charges.

In the $\mathbb{C}\mathrm{P}^1$ model the instantons and anti-instantons exhaust all possible static solutions. But in general the $\mathbb{C}\mathrm{P}^n$ models have static solutions that are neither instantons nor anti-instantons,[26] *i.e.* u^α is a function of both z and $\bar{z}$. Such solutions correspond to saddle points of the energy.

Finally we return to the Skyrme model in (3+1) dimensions. It has a very similar topology to the σ-models: finite energy again imposes a compactification of the spatial variables, this time into a three-sphere. The possible field configurations are maps from S^3 to S^3, and so are labelled by a winding number, usually called B (since from the point of view of particle physics it plays the role of the conserved baryon number). Explicitly

$$B = \frac{1}{24\pi^2}\int \epsilon_{ijk}\,\mathrm{Tr}\big((J^{-1}\partial_i J)(J^{-1}\partial_j J)(J^{-1}\partial_k J)\big)\,d^3x\,, \tag{2.63}$$

where ϵ_{ijk} is the alternating tensor in three dimensions. By completing the square, the potential energy in the Skyrme model may be cast in the form

$$E = -\tfrac{1}{4}\,\mathrm{Tr}\Big(J^{-1}\partial_i J \pm \tfrac{1}{4}\epsilon_{ijk}\,[J^{-1}\partial_j J\,,\,J^{-1}\partial_k J]\Big)^2 \pm 12\pi^2 B\,, \tag{2.64}$$

where the coupling constant e has been set equal to unity for convenience. The first term on the right hand side of (2.64) is always positive and so, just as before, there is a lower

bound on E:

$$E \geq 12\pi^2|B|\,, \tag{2.65}$$

with equality if and only if

$$J^{-1}\partial_i J \pm \tfrac{1}{4}\epsilon_{ijk}\,[J^{-1}\partial_j J\,,\; J^{-1}\partial_k J] = 0\,. \tag{2.66}$$

At this point an important distinction between the Skyrme model and the pure σ-models becomes apparent, because the Bogomolny equations (2.66) have no solution apart from the trivial one with zero energy, in which J is a constant. In other words the Bogomolny bound cannot be attained. In the language of condensed matter physics, the model is said to be "frustrated". The skyrmion configuration has no analytic expression; instead it must be computed numerically. It has been found that its energy exceeds the Bogomolny bound by about 23%.[27] (However, it is interesting to note that if the skyrmion is allowed to live in *curved* spacetime then the Bogomolny bound *can* be attained.[28])

To sum up, for models with nontrivial topology one can construct a lower bound on the potential energy in a given topological sector. It is usually proportional to the topological charge. Provided that the bound can be attained, the corresponding (static) solitons arise as solutions of the (first-order) Bogomolny equations. For instance, this is the situation in the (2+1)-dimensional Kähler models (where the static solitons are known as instantons). However, there are also so-called frustrated models, such as the Skyrme model, in which the bound cannot be attained. In these cases one must either make do with an approximate solution, or else resort to numerical computation.

Chapter III

Extended Waves in an Integrable Chiral Model

Many nonlinear field theories are said to be "integrable". Roughly speaking, this means that procedures exist for writing down their solutions explicitly. For example, many models (especially in (1+1) dimensions) are integrable via the inverse scattering method. More formally, integrable theories are characterized by the existence of an infinite number of conserved currents, by the existence of a Lax pair,[29] and by the Painlevé property.[30]

The σ-models are integrable in both (1+1)[31,32] and (2+0) dimensions. Note that solutions in (2+0) may be thought of as static configurations of the same theory in (2+1). We have already exhibited integrability in this case by constructing instanton solutions. On the other hand, the full (2+1)-dimensional σ-models are nonintegrable, *i.e.* introducing time dependence destroys the integrability.

From the physicist's point of view, it is interesting to have theories that are both integrable and Lorentz invariant. Some theories *are* fully integrable in (2+1) dimensions, for example the Davey-Stewartson equation[33,34] and the Kadomtsev-Petviashvili equation[35]; but both these are a long way from being Lorentz invariant, in contrast with the σ-models, which are Lorentz invariant but nonintegrable. It is not known whether it is possible for a theory to possess both properties. If such a model did exist then it would surely be of great interest in physics.

A partial remedy is to take an SO(2,1) invariant model and to modify it slightly in such a way as to trade Lorentz invariance for integrability. In particular, it is useful to look at modifications of the σ-models. One may still hope for a "generalized" Lorentz invariance in the sense that the behaviour of the soliton solutions, or of some restricted class of soliton solutions, is Lorentz invariant. An integrable modification of the SU(2) chiral model has been studied by Ward.[36] It was found that there are lump-like solitons, which scatter trivially off each other in the sense that each lump continues to move in a straight line with constant velocity. The aim of this chapter is to investigate wave-like solitons in the same model.

3.1 Introduction

The modified SU(2) chiral model studied by Ward is given by the field equation

$$(\eta^{\mu\nu} + V_\alpha \epsilon^{\alpha\mu\nu})\partial_\mu(J^{-1}\partial_\nu J) = 0. \tag{3.1}$$

As before, J takes values in SU(2) and is thought of as a 2×2 matrix of functions of the spacetime coordinates (t, x, y), sometimes also written (x^0, x^1, x^2). Greek letters are spacetime indices, taking values $0, 1, 2$, and ∂_μ denotes partial differentiation with respect to x^μ. The quantity $\epsilon^{\alpha\mu\nu}$ is the alternating tensor on three indices (with ϵ^{012} taken equal to $+1$) and $\eta^{\mu\nu} = \mathrm{diag}(-1, +1, +1)$ is the (inverse) Minkowski metric. Finally, V_α is a constant vector in spacetime.

Choosing $V_\alpha = (0, 0, 0)$ corresponds to the unmodified chiral model, which is Lorentz invariant but nonintegrable. Note that a nonzero V_α explicitly breaks Lorentz invariance by picking out a particular direction in spacetime. A case of particular interest occurs when V_α is chosen to be a spacelike unit vector, *i.e.* $V_\mu V^\mu = +1$, since then the theory appears to be integrable.[37] (Further motivation, in terms of reduction of the self-dual Yang-Mills equations, is given in appendix A.) Moreover, if $V_0 = 0$ then the theory possesses the same conserved energy-momentum vector as the unmodified chiral model, namely

$$P_\mu = (-\delta^\alpha_\mu \delta^\beta_0 + \tfrac{1}{2}\eta_{\mu 0}\eta^{\alpha\beta}) \,\mathrm{Tr}(J^{-1}J_\alpha J^{-1}J_\beta). \tag{3.2}$$

The corresponding energy density is

$$P_0 = -\tfrac{1}{2}\,\mathrm{Tr}\big((J^{-1}J_t)^2 + (J^{-1}J_x)^2 + (J^{-1}J_y)^2\big). \tag{3.3}$$

Here δ^α_μ is the Kronecker delta, Tr denotes the matrix trace and $J_\alpha \equiv \partial_\alpha J$. It should be emphasized that P_0 is a positive-definite functional of the field J.

If $V_0 \neq 0$ then it is not at all clear that a conserved energy-momentum vector exists and so from now on, in order to ensure integrability and a conserved energy, we shall take V_α to be a spacelike unit vector with $V_0 = 0$. To be specific, choose $V_\alpha = (0, 1, 0)$. Ward has shown that this model admits solitons, localized in two dimensions, which pass through each other without scattering or changing shape. It is the purpose of this chapter to construct extended plane wave solutions and to investigate their interactions.

Such waves are localized along the direction of motion, but have infinite spatial extent perpendicular to it.

In fact, one family of extended solutions may be exhibited immediately by noting that (3.1) is a generalization of the sine-Gordon (SG) equation in $(1+1)$ dimensions. Consider a J of the form

$$J = \begin{pmatrix} \cos\frac{1}{2}\phi & e^{-2ix}\sin\frac{1}{2}\phi \\ -e^{2ix}\sin\frac{1}{2}\phi & \cos\frac{1}{2}\phi \end{pmatrix}, \tag{3.4}$$

where the field ϕ depends on y and t, but not on x. The field equation (3.1) with $V_\alpha = (0,1,0)$ is then equivalent to the SG equation for ϕ:

$$\phi_{tt} - \phi_{yy} + 4\sin\phi = 0. \tag{3.5}$$

Furthermore, the energy density (3.3) becomes

$$P_0 = \tfrac{1}{4}(\phi_t^2 + \phi_y^2) + 4\sin^2\tfrac{1}{2}\phi\,, \tag{3.6}$$

which is precisely the energy density of the sine-Gordon theory. In other words, there are solutions which look like SG solitons living in the (1+1)-dimensional subspace spanned by (y,t), but spatially extended along the x-axis. Note that while the J of equation (3.4) depends explicitly on x, the corresponding energy density (3.6) does not. This illustrates a general feature of extended wave solutions: although P_0 depends only on time together with one spatial coordinate (along the direction of motion), J is necessarily a function of all three spacetime coordinates.

3.2 Construction of Solutions

This section summarizes the general method for constructing multisoliton solutions of the field equation (3.1). The technique is a variation of the well-known "Riemann problem with zeros" (see, for example, Forgács *et al.*[38]), and full details are to be found in the paper by Ward.[36]

There are two ingredients to an n-soliton solution. Firstly, a set of n complex numbers μ_k (k taking values from 1 to n), which must all be different and nonreal; secondly, for

each k, a meromorphic function f_k of the linear combination

$$\omega_k = x + \tfrac{1}{2}\mu_k(t+y) + \tfrac{1}{2}\mu_k^{-1}(t-y). \tag{3.7}$$

Now form the two-component objects $m_a^k = (1, f_k)$, so that a takes values $1, 2$ with $m_1^k = 1$ and $m_2^k = f_k$. Then (the inverse of) a matrix J that satisfies the field equation is given by

$$(J^{-1})_{ab} = \frac{1}{\sqrt{\alpha}}\Big(\delta_{ab} + \sum_{k,l}\frac{1}{\mu_k}(\Gamma^{-1})^{kl}\bar{m}_a^l m_b^k\Big), \tag{3.8}$$

where

$$\Gamma^{kl} = \sum_{a=1}^{2}(\bar{\mu}_k - \mu_l)^{-1}\bar{m}_a^k m_a^l,$$

$$\alpha = \prod_{k=1}^{n}\bar{\mu}_k/\mu_k$$

and bar denotes complex conjugation.

Clearly, the expression for J becomes very complicated very quickly as n is increased. Fortunately, there is plenty of analysis which can be done while still taking n small. For the rest of this section and the whole of the next, n will be equal to 1. Later on, a study of interactions (sections 3.4 and 3.5) will require $n = 2$.

To get a feel for the physical picture, we shall first investigate a simple family of lump solutions, very similar to those discussed by Ward. Consider $n = 1$, in which case solutions are specified by a complex number μ and a meromorphic function $f(\omega)$. Equation (3.8) simplifies to give

$$J = \frac{1}{|\mu|(1+|f|^2)}\begin{pmatrix}\mu + \bar{\mu}|f|^2 & (\mu-\bar{\mu})f\\ (\mu-\bar{\mu})\bar{f} & \bar{\mu}+\mu|f|^2\end{pmatrix}. \tag{3.9}$$

Writing $\mu = me^{i\theta}$, the energy density becomes

$$P_0 = \frac{2(1+m^2)^2\sin^2\theta}{m^2}\ \frac{|f'|^2}{(1+|f|^2)^2}, \tag{3.10}$$

where f' is the derivative of f as a function of ω. Keeping things simple, choose $f(\omega) = a\omega + c$ where $a \in \mathbb{R}$ and $c \in \mathbb{C}$. (One could generate a larger set of solutions by taking a

also in $\mathbb{C}$, but this is a little more tricky to analyse and an unnecessary complication to introduce at this stage.) The factor $|f'|^2$ in the numerator of (3.10) becomes just a^2. So it is seen that the solution looks like a single lump located at the point where $f = a\omega + c = 0$. From (3.7), its velocity is computed to be

$$(v_x, v_y) = \Big(\frac{-2m\cos\theta}{1+m^2}\,,\ \frac{1-m^2}{1+m^2}\Big). \tag{3.11}$$

The parameters μ, a and c have simple physical interpretations: μ specifies the soliton velocity via (3.11), c determines the position of the peak at time $t = 0$ and, finally, a fixes the ratio of the height of the lump to its width. Note that in the static case ($\mu = i$) one may easily integrate P_0 over x and y to obtain the total energy E. The result is

$$E \equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_0\,\mathrm{d}x\,\mathrm{d}y = 8\pi\ , \tag{3.12}$$

which is independent of a.

3.3 Extended Wave Solutions

Now we shall set out to construct a family of extended wave solutions. Ward showed that taking f to be rational of degree N leads to a configuration with N peaks, which in the static case has energy $8N\pi$. An extended wave must have infinite energy and so no function of finite degree will do for f. The next candidate is some sort of exponential. Specifically, consider

$$f(\omega) = \exp(b\omega + c). \tag{3.13}$$

This leads to an energy density

$$P_0 \;=\; \frac{2(1+m^2)^2\sin^2\theta}{m^2}\ \ \frac{|b|^2|f|^2}{(1+|f|^2)^2}\ . \tag{3.14}$$

Here $\mu = me^{i\theta}$ as before. Note that P_0 depends on c only through its real part and so, without loss of generality, we can take $c \in \mathbb{R}$. However, b is in general complex. To see

that (3.14) does indeed look like a wave, rewrite it as

$$P_0 = \frac{(1+m^2)^2 \sin^2\theta}{2m^2} |b|^2 \operatorname{sech}^2\big(\mathrm{Re}(b\omega) + c\big). \tag{3.15}$$

Note that P_0 is constant along each of the lines $\mathrm{Re}(b\omega) + c = \text{const}$. The wavefront (*i.e.* the crest of the wave) lies along $\mathrm{Re}(b\omega) + c = 0$. For each value of t, this is the equation of a straight line in the xy-plane. As t varies, the wave maintains its shape and simply moves at constant velocity.

To investigate this wave in more detail, write $b = |b|e^{i\alpha}$. Then the equation of the wavefront may be written

$$Ax + By = Ct + D\,, \tag{3.16}$$

where

$$\begin{aligned} A &= 2m\cos\alpha\,, \\ B &= m^2\cos(\theta+\alpha) - \cos(\theta-\alpha)\,, \\ C &= -m^2\cos(\theta+\alpha) - \cos(\theta-\alpha)\,, \\ D &= -2mc/|b|\,. \end{aligned}$$

The velocity may be readily calculated:

$$\begin{aligned} v_x &= \frac{-2m\cos\alpha\big(\cos(\theta-\alpha) + m^2\cos(\theta+\alpha)\big)}{m^4\cos^2(\theta+\alpha) + 2m^2(\sin^2\theta + \cos^2\alpha) + \cos^2(\theta-\alpha)}\,, \\ v_y &= \frac{\cos^2(\theta-\alpha) - m^4\cos^2(\theta+\alpha)}{m^4\cos^2(\theta+\alpha) + 2m^2(\sin^2\theta + \cos^2\alpha) + \cos^2(\theta-\alpha)}\,. \end{aligned} \tag{3.17}$$

The speed v is given by

$$v^2 = 1 - \frac{4m^2\sin^2\theta}{m^4\cos^2(\theta+\alpha) + 2m^2(\sin^2\theta + \cos^2\alpha) + \cos^2(\theta-\alpha)} \tag{3.18}$$

and, although the integral of P_0 over all space is divergent, one can instead calculate the energy per unit length along the wavefront, which turns out to be

$$\widehat{E} = \frac{4\gamma\,|b|\,|\sin^3\theta|\,(1+m^2)^2}{m^4\cos^2(\theta+\alpha) + 2m^2(\sin^2\theta + \cos^2\alpha) + \cos^2(\theta-\alpha)}\,, \tag{3.19}$$

where $\gamma = (1-v^2)^{-1/2}$.

The question of classification of these solutions now arises. Clearly, c determines the position of the wave at time $t = 0$. On the face of it, there are four other real parameters (m, θ, $|b|$ and α), which one might naïvely think could be chosen to fix the velocity (two parameters), and the wave height and the wave width (one parameter each) all independently. If this were the case, it would support the conjecture of generalized Lorentz invariance. However, the following systematic study of some special cases shows that things are not quite so simple.

It is not difficult to pick out the solutions which look like static waves aligned along the coordinate axes. A wave lying on the x-axis requires $\cos\alpha = 0$ and $m = 1$. Setting $k = ib\sin\theta$ leads to

$$P_0 = 2k^2 \operatorname{sech}^2 ky \,, \tag{3.20}$$

which is (if one parametrizes the solution using k and θ, rather than $|b|$ and θ) independent of θ. On the other hand, the conditions for a wave to lie on the y-axis are $\cos\theta = \sin\alpha = 0$. Setting $k = |b|$ leads to

$$P_0 = \frac{(1+m^2)^2}{2m^2} k^2 \operatorname{sech}^2 kx \,. \tag{3.21}$$

In the latter case the height and width may be chosen independently, while in the former they are determined by a single parameter, with θ playing the role of an "internal" degree of freedom. Although these observations do not necessarily rule out a generalized Lorentz invariance (there may be other plane wave solutions, not generated by (3.13)), they make it seem unlikely.

The complete classification of waves generated by (3.13) appears to be difficult. Therefore it is useful to study subsets of solutions obtained by imposing some extra condition on the parameters. For example, if one requires

$$(1+m^2)\tan\alpha = (1-m^2)\tan\theta \,, \tag{3.22}$$

then the wave velocity (3.17) becomes

$$(v_x, v_y) = \left(-\frac{2m\cos\theta}{1+m^2} \,,\, \frac{1-m^2}{1+m^2}\right) , \tag{3.23}$$

which matches the expression (3.11) for lump solutions, and the energy per unit length

becomes

$$\widehat{E} = 4\gamma\,|b|\,|\sin^3\theta|. \tag{3.24}$$

In this scheme a physical interpretation becomes apparent: m and θ specify the velocity and $|b|$ determines both the height and width, with α fixed by (3.22). To make the situation even more transparent, one can replace $(m, |b|, \theta)$ with new parameters (k, ϕ, A) defined by

$$\begin{aligned} m &= \left(\frac{1 - k\sin\phi}{1 + k\sin\phi}\right)^{\frac{1}{2}}, \\ \cos\theta &= \frac{-k\cos\phi}{\sqrt{1 - k^2\sin^2\phi}}, \\ |b| &= A\sqrt{1 - k^2\sin^2\phi}\,. \end{aligned} \tag{3.25}$$

The energy density then becomes

$$P_0 = \frac{2A^2(1-k^2)}{1 - k^2\sin^2\phi}\ \mathrm{sech}^2\, A(x\cos\phi + y\sin\phi - kt). \tag{3.26}$$

Now it is seen that k is the wave speed, ϕ is the angle of the direction of motion relative to the x-axis and A is the width. The height is then a simple function of k, ϕ and A. Taking $\phi = \pi/2$ and $A = 2/\sqrt{1-k^2}$ leads to the extended SG waves mentioned earlier, namely

$$P_0 = \left(\frac{8}{1-k^2}\right)\ \mathrm{sech}^2\,\frac{2(y-kt)}{\sqrt{1-k^2}}\,. \tag{3.27}$$

Note that in this case, equation (3.24) reduces to $\widehat{E} = 8\gamma$, so confirming the relativistic behaviour of SG solitons.

3.4 Wave-Wave Interactions

It was seen in the previous section that the classification of extended wave solutions is far from a trivial matter, and clearly a complete study of their interactions will be no simpler. Instead, we shall present an analysis of a few particular cases, pointing out the main features. It seems likely that the general case will be very similar.

For a two-soliton solution one takes $n = 2$ in the prescription of section 3.2. It turns out that the algebra is much simplified if μ_k is restricted to be pure imaginary, *i.e.* $\theta_k = \pi/2$. So as an example of a solution containing two waves, W_1 and W_2, consider

$$\mu_k = ip_k\,, \qquad f_k(\omega_k) = \exp(b_k\omega_k + c_k)\,, \tag{3.28}$$

where k takes values $1, 2$; p_k is real and

$$b_k = A_k\big((1+p_k^2)\cos\phi_k - 2ip_k\sin\phi_k\big)\,. \tag{3.29}$$

Physically, the positive real parameter A_k fixes the width and height of each wave, $\phi_k \in [0,\pi]$ gives the direction of motion and the speed is

$$\sin\phi_k\left(\frac{1-p_k^2}{1+p_k^2}\right).$$

Even with the simplification of taking μ_k imaginary, the full expression for P_0 is rather complicated, but one can investigate the asymptotic behaviour in the following sense. Recall that the equation of each wavefront is $\mathrm{Re}(b_k\omega_k) + c_k = 0$. Taking the limits $\mathrm{Re}(b_1\omega_1) \to \pm\infty$ corresponds to moving far away from wave W_1 on either side. If at the same time $\mathrm{Re}(b_2\omega_2)$ is kept finite, then roughly speaking we are keeping our eyes fixed on W_2, but far away from W_1. To keep things in terms of f_k, note that $\mathrm{Re}(b_k\omega_k) \to +\infty$ implies $|f_k| \to \infty$ and $\mathrm{Re}(b_k\omega_k) \to -\infty$ implies $|f_k| \to 0$.

Now let k' stand for "not k", so that $1' = 2$ and $2' = 1$. Then the asymptotics of the solution (3.28) may be summarized as follows:

$$|f_{k'}| \to \infty\,, \qquad P_0 \sim \frac{2(p_1^2-p_2^2)^2(p_k^2+1)^2A_k^2\big((p_k^2+1)^2-(p_k^2-1)^2\sin^2\phi_k\big)|f_k|^2}{p_k^2\big(|f_k|^2(p_1-p_2)^2+(p_1+p_2)^2\big)^2}\,,$$

$$|f_{k'}| \to 0\,, \qquad P_0 \sim \frac{2(p_1^2-p_2^2)^2(p_k^2+1)^2A_k^2\big((p_k^2+1)^2-(p_k^2-1)^2\sin^2\phi_k\big)|f_k|^2}{p_k^2\big(|f_k|^2(p_1+p_2)^2+(p_1-p_2)^2\big)^2}\,. \tag{3.30}$$

The crucial point is the difference of sign in the denominators. It is not difficult to see

that the essential behaviour is

$$\begin{aligned} |f_{k'}| \to \infty \ , \qquad & P_0 \sim \operatorname{sech}^2\big(\mathrm{Re}(b_k\omega_k) + c_k - \gamma\big) \ , \\ |f_{k'}| \to 0 \ , \qquad & P_0 \sim \operatorname{sech}^2\big(\mathrm{Re}(b_k\omega_k) + c_k + \gamma\big) \ , \end{aligned} \tag{3.31}$$

where

$$\tanh\gamma = \frac{2p_1p_2}{p_1^2 + p_2^2} \ . \tag{3.32}$$

So the waves interact in a fairly simple way: each experiences a phase shift 2γ. The SG waves are present in the above solutions as the special case $\phi_k = \pi/2$, $A_k = 1/|p_k|$.

Figure 3.1 shows a snapshot of the energy density at time $t = 0$ for the following choice of parameters: $(p_1, A_1, \phi_1) = (1, 2, \pi/2)$, $(p_2, A_2, \phi_2) = (2, 1, \pi/4)$. The phase shift suffered by each wave is clearly visible. Note also the highly nonlinear superposition in the region of intersection.

One may ask whether internal parameters (which do not appear in the single wave energy density) can affect interactions. The answer is yes, as the following example will show. Consider W_1 and W_2 both parallel to the x-axis with W_2 stationary, *i.e.* $\phi_1 = \phi_2 = \pi/2$, $p_2 = 1$. Choose $p_1 = 1/\sqrt{2}$. Then in the above scheme, $\tanh\gamma = 2\sqrt{2}/3$. But now note that W_2 is equally well described by

$$\mu_2 = \exp(i\pi/4) \ , \qquad f_2(\omega_2) = \exp(-2\sqrt{2}iA_2\omega_2 + c_2) \ .$$

Repeating the calculation (although μ_2 is not now pure imaginary, the parameters have been chosen to make the algebra as tractable as possible) one finds precisely the same asymptotic behaviour but with a new phase shift γ' given by $\tanh\gamma' = 2/3$.

To sum up, as two waves interact, they do not change shape or velocity, but each has a phase shift across the region of intersection, which may be dependent upon internal parameters.

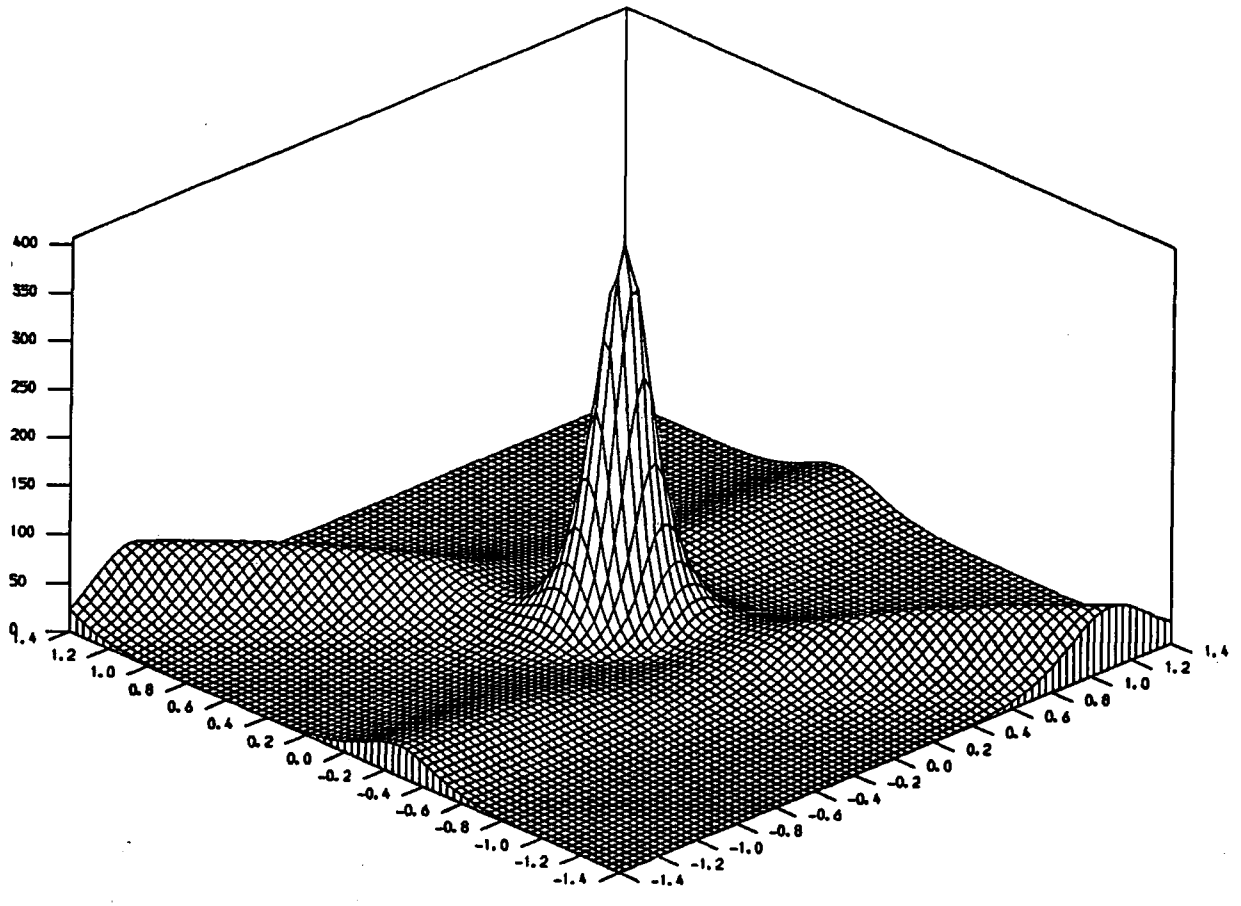


FIG 3.1. A snapshot of the energy density for a two-wave interaction. The flatter wave is stationary along the x-axis, and the taller one is moving across it at an angle of 45 degrees.

3.5 Wave-Lump Interactions

For simplicity, we shall consider only the case of a plane wave W_1 incident on a stationary lump L_2. In terms of the input to the two-soliton solution, take the same μ_1 and f_1 as in section 3.4, but now with $\mu_2 = i$ and $f_2(\omega_2) = A_2\omega_2$ ($A_2 \in \mathbb{R}$ as in section 3.2). The effects of the interaction are again revealed by the asymptotic behaviour of P_0. Note that to look far away from L_2 in any direction, the relevant limit is $|f_2| \to \infty$. One

finds the following:

$$|f_1| \to \infty \ , \qquad P_0 \sim \frac{8(p_1^2-1)^2 A_2^2}{\left(|f_2|^2(p_1-1)^2+(p_1+1)^2\right)^2} \ ,$$

$$|f_1| \to 0 \ , \qquad P_0 \sim \frac{8(p_1^2-1)^2 A_2^2}{\left(|f_2|^2(p_1+1)^2+(p_1-1)^2\right)^2} \ , \tag{3.33}$$

$$|f_2| \to \infty \ , \qquad P_0 \sim \frac{2A_1^2(p_1^4-1)^2\left((p_1^2+1)^2-(p_1^2-1)^2\sin^2\phi_1\right)|f_1|^2}{p_1^2\left(|f_1|^2(p_1-1)^2+(p_1+1)^2\right)^2} \ .$$

The physical picture is this: the shape and velocity of the wave are the same long before and long after the collision, and it suffers no phase shift. The more remarkable feature is that the lump remains stationary, but changes its height by a factor

$$\left(\frac{p_1-1}{p_1+1}\right)^4 .$$

Again the crucial point is the difference of sign in the denominators. A little care is needed at this stage, since it is not immediately clear which limit of $|f_1|$ corresponds to $t \to -\infty$ and which to $t \to +\infty$. The answer to this question depends on the size of p_1:

$$\text{For } |p_1|<1, \qquad \begin{cases} t \to -\infty & \Rightarrow \quad |f_1| \to \infty \ . \\ t \to +\infty & \Rightarrow \quad |f_1| \to 0 \ . \end{cases}$$

$$\text{For } |p_1|>1, \qquad \begin{cases} t \to -\infty & \Rightarrow \quad |f_1| \to 0 \ . \\ t \to +\infty & \Rightarrow \quad |f_1| \to \infty \ . \end{cases}$$

So for $p_1 > 1$ or $-1 < p_1 < 0$ the lump decreases in height and for $p_1 < -1$ or $0 < p_1 < 1$ it increases in height. Figure 3.2 shows a series of snapshots taken at time intervals of 0.5, starting at $t = -1.0$, for the following parameters: $\phi_1 = \pi/2$, $A_1 = 0.1$, $A_2 = 5$, $p_1 = 10$. In this case, W_1 is an SG wave. The lump decreases in height by a factor $(9/11)^4$ (≈ 0.45), but its total energy remains unchanged, equal to 8π.

Perhaps the most puzzling feature is the transverse asymmetric kink acquired by the wave as it squashes the lump, and which then gradually dies away. It could be that this is due somehow to the absence of Lorentz invariance. Alternatively, internal parameters at work may provide the explanation. In any event, the interaction seems quite unlike any occurring in other integrable models.

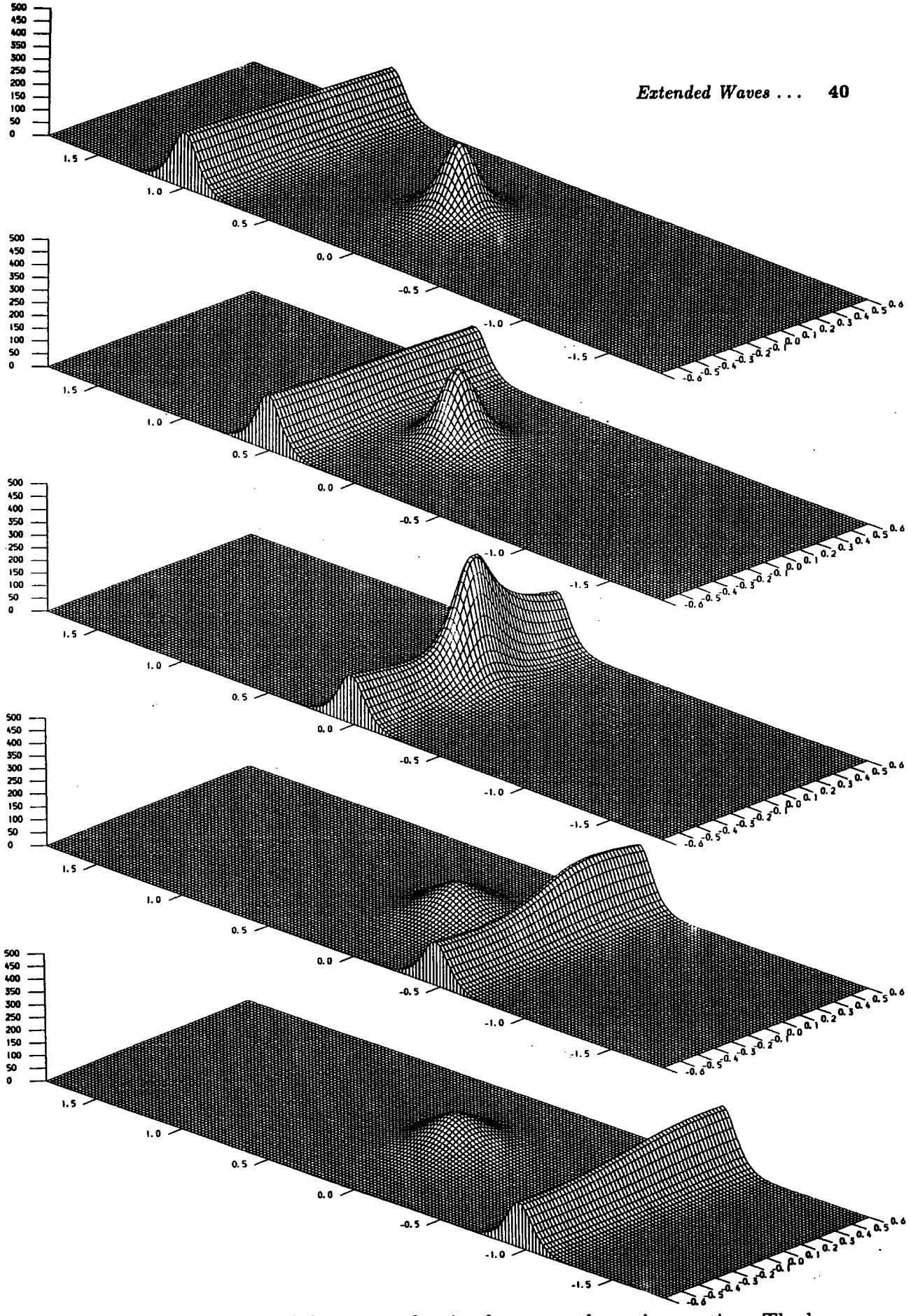


FIG 3.2. A series of snapshots of the energy density for a wave-lump interaction. The lump is stationary at the origin and the wave is travelling parallel to the y-axis. Time runs down the page in intervals of 0.5, starting at $t = -1.0$.

3.6 Concluding Remarks

The modified SU(2) chiral model and the Kadomtsev-Petviashvili (KP) equation have several features in common. The latter also possesses "rational" solitons, which look like lumps, and "exponential" solitons, which look like waves.[39] In both models two lumps pass through each other without scattering and two waves interact with a phase shift. However, the wave-lump interaction of section 3.5 seems to have no analogue in KP; compare, for example, with figure 9 of reference [39].

As a final remark, it might be interesting to consider letting the field J live in a noncompact Lie group such as $SL(2,\mathbb{R})$ or $SL(2,\mathbb{C})$. This would mean that the energy density is no longer positive definite, but should not rule out explicit construction of solutions. An $SL(2,\mathbb{R})$ model is expected to have embedded in it the Korteweg-de Vries (KdV) equation in (1+1) dimensions, while taking J in $SL(2,\mathbb{C})$ will also include the nonlinear Schrödinger equation.[40] Maybe these models will exhibit a behaviour closer to KP, since they are both, in some sense, generalizations of KdV, unlike the current SU(2) model.

Chapter IV

Discrete Bogomolny Equations in the O(3) Model

This chapter and the next deal with the O(3) σ-model, and in particular with an investigation of soliton stability. Since the solitons have no fixed size, their stability is a central question. Nontrivial topology guarantees that their potential energy can never fall below the Bogomolny bound; but there is still the possibility that under small perturbations they could shrink towards infinitely tall spikes of zero width.

Since the model is nonintegrable, it is natural to consider numerical techniques, in which the field equations are discretized and then evolved on a lattice. A lattice study of soliton stability can be hampered by the absence of explicit static solutions. This chapter will show how to avoid this problem by taking proper account of the model's topological aspects. The application of the techniques developed here to the question of soliton stability is described in the next chapter.

At first sight it is not clear how best to reconcile the nontrivial topological aspects with a lattice formulation, although there have been several proposals.[41–44] The approach adopted here is to construct a discrete analogue of the Bogomolny bound. This leads naturally to a numerical scheme that possesses explicit static solutions. In other words, the central idea is to find a set of first-order difference relations to play the role of the Bogomolny equations on a lattice. Solutions of these relations saturate the topological lower bound on the (discretized) energy.

Only field configurations for which the energy density (but not necessarily the fields themselves) is radially symmetric will be considered here. In principle this restriction could be lifted, but the construction of the discrete Bogomolny relations is then more complicated. However, for the problem of soliton stability, the restriction to axial symmetry is not a serious constraint, since nonaxial modes are unlikely to lead to instabilities. It is worth remarking that Mikhailov and Yaremchuk have considered imposing radial symmetry on the fields themselves.[45] Then the model is integrable via an inverse scattering method, but all the solutions obtained in this way have topological charge zero.

The rest of this chapter is arranged as follows. The next section reparametrizes the

fields to impose radial symmetry. Only then is the model discretized, and in section 4.2 the discrete Bogomolny relations are constructed. Section 4.3 discusses their properties (with some of the more mathematical details gathered in appendix B). In sections 4.4 and 4.5 they are incorporated into a full evolution scheme (a program listing is given is appendix C). Some preliminary results, indicating that the numerical simulation is working well, are presented in section 4.6. In keeping with convention, the minimum energy static solutions (*i.e.* those solutions that attain the Bogomolny bound) will often be referred to as instantons.

4.1 Radial Symmetry in the O(3) Sigma Model

In the O(3) σ-model the fields $\boldsymbol{\phi}$ take values on a two-sphere S of unit radius. It is often useful to relate $\boldsymbol{\phi}$ to a complex scalar field u by means of a stereographic projection from the north pole of S on to the complex u-plane:

$$u = \frac{\phi_1 + i\phi_2}{1 - \phi_3} \,. \tag{4.1}$$

The radially symmetric instantons of charge N are given by $u = \lambda/z^N$ where $z = x + iy$ and λ is a real constant. Note that the global O(3) invariance of the model has now been removed by implicitly choosing $u \to \infty$ (equivalently $\boldsymbol{\phi} \to (0,0,1)$) as $z \to 0$, and by taking λ real. When $N = 0$ the field is constant and the energy density is zero everywhere. For $N = 1$ the instanton looks like a lump peaked at the origin, and for $N > 1$ it is a ring centred on the origin and peaked at

$$r = \left(\frac{N-1}{N+1}\right)^{1/2N} \lambda^{1/N} \,, \tag{4.2}$$

where r is the polar radius in the xy-plane. Physically, λ may be interpreted as the instanton width.

With these observations in mind, one may write down a more general family of radially symmetric configurations, namely

$$u = \frac{\lambda(r,t)}{z^N} \,, \tag{4.3}$$

where now λ is allowed to be a (possibly complex) function of r and t. The remainder of this chapter deals exclusively with these configurations. It seems that all radially

symmetric energy densities can be derived from a field of the form (4.3), at least in some global gauge.

The field u provides a concise description of the radial configurations. However, it is not suited to numerical implementation, since it possesses a singularity (at the origin). Instead, one reverts to the ϕ-picture, in which (4.3) is equivalent to

$$\begin{aligned} \phi_1 &= f(r,t)\cos N\theta + g(r,t)\sin N\theta\,, \\ \phi_2 &= -f(r,t)\sin N\theta + g(r,t)\cos N\theta\,, \\ \phi_3 &= h(r,t)\,, \end{aligned} \tag{4.4}$$

where f, g and h satisfy the constraint

$$f^2 + g^2 + h^2 = 1\,, \tag{4.5}$$

and θ is the polar angle in the xy-plane. Roughly speaking, f comes from the real part of λ and g from the imaginary part (so the instantons have $g = 0$). The boundary conditions are $h(0) = 1$ and, for $N \neq 0$, $h(\infty) = -1$.

Now one is in a position to reformulate the Bogomolny bound in terms of the single spatial coordinate r. Set $g \equiv 0$ in (4.4), since for the moment we are concerned only with instanton solutions, and choose f to be positive. Then the kinetic and potential energies are given by

$$T = \frac{1}{4}\int_0^\infty \left(\frac{h_t^2}{1-h^2}\right)(2\pi r\,dr)\,, \tag{4.6}$$

$$V = \frac{1}{4}\int_0^\infty \left(\frac{h_r^2}{1-h^2} + \frac{N^2(1-h^2)}{r^2}\right)(2\pi r\,dr)\,, \tag{4.7}$$

where the subscripts r and t denote partial differentiation. The topological charge density becomes simply

$$\rho \equiv -\frac{1}{8\pi}\epsilon_{ij}\boldsymbol{\phi}\,.\,(\partial_i\boldsymbol{\phi}\times\partial_j\boldsymbol{\phi}) = -\frac{N h_r}{4\pi r}\,. \tag{4.8}$$

The Bogomolny bound, $V \geq 2\pi N$, is essentially the identity

$$\frac{1}{4}\int_0^\infty \left(\frac{h_r}{\sqrt{1-h^2}} + \frac{N\sqrt{1-h^2}}{r}\right)^2 (2\pi r\,dr) \geq 0 \tag{4.9}$$

and the Bogomolny equations are

$$rh_r + N(1 - h^2) = 0\,. \tag{4.10}$$

Explicitly, the instanton solutions are given by

$$(f, g, h) = \left(\frac{2\lambda r^N}{\lambda^2 + r^{2N}}\,,\, 0\,,\, \frac{\lambda^2 - r^{2N}}{\lambda^2 + r^{2N}} \right)\,, \tag{4.11}$$

for arbitrary real constant λ. Note that (provided $N \neq 0$) h decreases monotonically as a function of r, from $h = 1$ at $r = 0$, to $h = -1$ as $r \to \infty$.

4.2 Bogomolny Relations in a Discrete Formulation

So far, all we have done is to re-express the (radially symmetric) instanton solutions in terms of a single real field h, which is a function only of the polar radius r. It will now be seen how this description is useful in constructing discrete analogues of the topological charge and of the Bogomolny equations.

Consider a discrete set of values h_n ($n \in \mathbb{Z}$, $n \geq 0$) with the properties that $h_0 = 1$ and $h_n \to -1$ as $n \to \infty$. (In both this section and the next, the special case $N = 0$ will be avoided, in order to ensure $h_n \to -1$.) One expects the Bogomolny bound to take the general form

$$\sum_{n=0}^{\infty} (\alpha_n + \beta_n)^2 = V - 2\pi N \geq 0\,. \tag{4.12}$$

By analogy with equation (4.9), the cross terms of the infinite sum in (4.12) should yield the topological charge; the remaining terms give the potential energy. So to fix up the charge one may take

$$\alpha_n \beta_n = \tfrac{1}{2}\pi N (h_{n+1} - h_n)\,. \tag{4.13}$$

Turning to the energy, the form of (4.7) suggests the choices

$$\alpha_n = \sqrt{\frac{\pi}{2}}\, N\, \frac{\sqrt{1 - h_n^2}}{\sqrt{n}}\,, \tag{4.14}$$

$$\beta_n = \sqrt{\frac{\pi n}{2}}\, \frac{(h_{n+1} - h_n)}{\sqrt{1 - h_n^2}}\,. \tag{4.15}$$

The only problem with this is that (4.14) and (4.15) are undefined when $n = 0$. Clearly the origin must be treated in a special way. One solution is to arrange that $\alpha_0 + \beta_0 = 0$ identically, while still being consistent with (4.13). So choose

$$\alpha_0 = \sqrt{\frac{\pi N}{2}}\sqrt{1 - h_1}\,, \tag{4.16}$$

$$\beta_0 = -\sqrt{\frac{\pi N}{2}}\sqrt{1 - h_1}\,. \tag{4.17}$$

Putting the pieces together gives the following discretized potential energy:

$$V_{\rm dis} = \pi N(1 - h_1) + \tfrac{1}{4}\sum_{n=1}^{\infty}(2\pi n)\left(\frac{(h_{n+1} - h_n)^2}{1 - h_n^2} + \frac{N^2(1 - h_n^2)}{n^2}\right). \tag{4.18}$$

Apart from the leading term, which comes from $\alpha_0^2 + \beta_0^2$, this is perhaps what one would have written down immediately as an analogue of (4.7). The advantage of the above approach lies in the appearance of the associated Bogomolny relations. There is equality in (4.12) if and only if $\alpha_n + \beta_n = 0$ for $n \geq 1$. (Recall that $\alpha_0 + \beta_0 = 0$ identically.) Substituting for α_n and β_n from (4.14) and (4.15) one finds

$$N(1 - h_n^2) + n(h_{n+1} - h_n) = 0\,. \tag{4.19}$$

This equation is the key to our lattice formulation. The proposal is that instantons on the lattice should satisfy (4.19). This discrete Bogomolny relation is a nonlinear first-order difference equation for h_n, and so its solutions contain one degree of freedom, which specifies the instanton size in some way. It is simplest to think of h_1 as the free parameter: roughly speaking, the closer h_1 is chosen to 1 then the larger the width of the corresponding instanton. One should also be clear about the role of N. There is no equation giving the topological charge in terms of a specified set of h_n. Rather, N is now also a (positive integer) parameter to be specified.

Given h_1 and N, it is clearly very simple to generate all other h_n by repeated application of (4.19). So far, (4.19) has resisted all attempts to write down the general solution in a closed form. But despite this one can still make considerable progress in investigating the properties of lattice instantons: this is the subject of the next section.

As a final remark, one might ask how things would differ if the restriction of radial symmetry were dropped. The answer is that one would now need two scalar fields, labelled with two indices each. Moreover, to get the topological charge appearing in the Bogomolny bound correctly, the right hand side of (4.13) would look like the area of a spherical triangle (see, for example, the paper by Berg and Lüscher[41]).

4.3 Instantons on the Lattice

All lattice instantons satisfy the Bogomolny relations, but does the converse hold, *i.e.* are all solutions of (4.19) lattice instantons? The answer is no: the requirement that $h_n > -1$ for all n puts a lower bound on the allowed values of h_1, as the following argument shows. From (4.19)

$$h_{n+1} = h_n - \frac{N}{n}(1 - h_n^2) \tag{4.20}$$

and so, assuming $h_n > -1$,

$$\begin{aligned} h_{n+1} > -1 &\Leftrightarrow \; h_n + 1 > \frac{N}{n}(1 - h_n^2) \\ &\Leftrightarrow \; 1 > \frac{N}{n}(1 - h_n) \\ &\Leftrightarrow \; h_n > 1 - \frac{n}{N}\,. \end{aligned} \tag{4.21}$$

The condition (4.21) is automatically satisfied if $n \geq 2N$, but for $n < 2N$ it gives a set of $2N - 1$ inequalities, which are equivalent to putting a lower bound on h_1. When $N = 1$ one requires simply that $h_1 > 0$ (and, of course, $h_1 < 1$). For $N = 2$, (4.21) becomes

$$h_1 > \tfrac{1}{2}, \qquad h_2 > 0, \qquad h_3 > -\tfrac{1}{2}\,. \tag{4.22}$$

Using the explicit Bogomolny relations, namely

$$\begin{aligned} h_2 &= 2h_1^2 + h_1 - 2\,, \\ h_3 &= h_2^2 + h_2 - 1\,, \end{aligned} \tag{4.23}$$

one finds that the three conditions in (4.22) are all satisfied if and only if $h_1 > \sqrt{3}/2$.

These results are a little curious. One possible interpretation is that the lattice "wants" to support only those instantons larger than a certain width. This is somehow in keeping with the intuitive notion that a good discrete representation will put several lattice points inside the instanton. From now on it is assumed that h_1 is always large enough to satisfy (4.21).

One further check must be made. If $|h_n| < 1$ for some particular n then from (4.20) $h_{n+1} < h_n < 1$. Hence, subject to the above provisos on h_1, $\{h_n\}$ is a monotone decreasing sequence bounded below by -1, and so must tend to some limit, l, where $l \geq -1$. For an instanton solution, one requires $l = -1$, and it is easy to show that this is indeed always the case: since $\{h_n\}$ tends to a limit, the series

$$\sum_{n=1}^{\infty}(h_n - h_{n+1})$$

converges. By (4.20) this series is equal to

$$\sum_{n=1}^{\infty}\frac{N}{n}(1 - h_n^2)\,, \tag{4.24}$$

and if $l > -1$ then (4.24) diverges. So $l = -1$.

One could argue that the classification of lattice instantons is now complete. However, h_1 is not a convenient parameter to work with in practice. It would be much nicer to have a lattice analogue of the instanton width λ, which appeared in (4.11). One possibility is to set

$$\lambda_n^2 \equiv \frac{1 + h_n}{1 - h_n}\,n^{2N} \tag{4.25}$$

and then to define

$$\lambda \equiv \lim_{n\to\infty}\lambda_n\,, \tag{4.26}$$

provided this limit exists. The idea is that one could specify λ and somehow relate it back to h_1. It turns out that λ is indeed well defined: the proof of this is given in appendix B. However, it seems difficult to relate λ to h_1. Instead, it may be related to h_{n_0} for large

n_0:

$$\lambda^2 = \frac{(n_0+2N-1)!}{(n_0-1)!}\left(\frac{1+h_{n_0}}{1-h_{n_0}}\right)\exp(-4N^2S_0) + \mathcal{O}(n_0^{-3}), \tag{4.27}$$

where

$$S_0 \equiv \frac{\pi^2}{6} - \sum_{n=1}^{n_0-1}\frac{1}{n^2}. \tag{4.28}$$

This result is also derived in appendix B. It suggests the following procedure for constructing instantons on the lattice:

1. Specify λ and choose n_0 large enough so that $\mathcal{O}(n_0^{-3})$ may be neglected.
2. Calculate h_{n_0} using (4.27).
3. Calculate h_n for $n > n_0$ using (4.20).
4. Calculate h_n for $0 < n < n_0$ by solving (4.20) as a quadratic for h_n:

$$h_n = \frac{n}{2N}\left(-1+\sqrt{1+\frac{4N}{n}\left(\frac{N}{n}+h_{n+1}\right)}\right). \tag{4.29}$$

Consider (4.29) for a moment. As h_{n+1} varies between -1 and 1 so does h_n, and if $|h_{n+1}| < 1$ then $h_n > h_{n+1}$, *i.e.* step 4 is perfectly well behaved. The only problem will occur in step 3 if $n_0 < 2N$ and the value of h_{n_0} causes (4.21) to be violated (in which case the derivation of (4.27) breaks down anyway). So to be completely safe, one should always choose $n_0 > 2N$.

To sum up, lattice instantons may be generated by choosing h_1 and then repeatedly using (4.20), but there is a lower bound on the allowed h_1, which is dependent on N. It is much better to fix h_{n_0} for some large n_0; first because one need not worry about the allowed values of h_{n_0}, and secondly because h_{n_0} has been related, albeit approximately, to the width λ.

4.4 The Full Discrete Evolution Scheme

We shall now address the question of incorporating the notion of lattice instantons into a full numerical evolution scheme. One possibility is to write down the general time-dependent equations of motion for the continuum model and then to discretize them in some way, but the lattice instantons of section 4.3 will not in general be static solutions of these full discretized equations.

A better plan is to construct a discrete action in which the potential energy (4.18) of the lattice instantons appears, and then to vary the action with respect to the fields at each site on the spacetime lattice. This method guarantees that solutions of the Bogomolny relations (*i.e.* the lattice instantons) are automatically solutions of the full evolution scheme.

In the continuum model the kinetic and potential energy densities, ϵ_T and ϵ_V, are given in terms of f, g and h (defined in (4.4)) by

$$\begin{aligned} \epsilon_T(r,t) &= \frac{1}{4}(f_t^2 + g_t^2 + h_t^2)\,, \\ \epsilon_V(r,t) &= \frac{1}{4}\Big(f_r^2 + g_r^2 + h_r^2 + \frac{N^2}{r^2}(f^2+g^2)\Big)\,. \end{aligned} \tag{4.30}$$

It is now convenient to write

$$f(r,t) + ig(r,t) = R(r,t)\, e^{i\psi(r,t)}\,, \tag{4.31}$$

so replacing f and g with R and ψ. (Note the instantons have $g = 0$, *i.e.* $R = f$, $\psi = 0$.) We shall always choose $R > 0$. The constraint on the fields becomes $R^2 + h^2 = 1$, and so it is straightforward to eliminate R in favour of h, eventually having a theory containing just h and ψ. From (4.30) the action density may be written

$$\begin{aligned} \mathcal{L} &= \frac{1}{4}\Big(h_r^2 + R_r^2 + \frac{N^2R^2}{r^2} + R^2\psi_r^2 - h_t^2 - R_t^2 - R^2\psi_t^2\Big) \\ &= \frac{1}{4}\Big(\frac{h_r^2}{1-h^2} + \frac{N^2(1-h^2)}{r^2} + (1-h^2)(\psi_r^2 - \psi_t^2) - \frac{h_t^2}{1-h^2}\Big)\,. \end{aligned} \tag{4.32}$$

The first two terms are just the potential energy density (4.7) of instanton solutions; they will be discretized according to (4.18). In the full lattice formulation, the fields

acquire a superscript to label the time slices. The discrete action derived from (4.32) is

$$S_{\rm dis} = \pi N \sum_m (1 - h_1^m) + \tfrac{1}{2}\pi \sum_m \sum_{n=1}^{\infty} n \Big(\frac{(h_{n+1}^m - h_n^m)^2}{1 - h_n^{m2}} + \frac{N^2(1 - h_n^{m2})}{n^2} \\ - \frac{(h_n^{m+1} - h_n^m)^2}{\delta^2(1 - h_n^{m2})} + (1 - h_n^{m2}) \Big((\psi_{n+1}^m - \psi_n^m)^2 - \frac{1}{\delta^2}(\psi_n^{m+1} - \psi_n^m)^2 \Big) \Big) , \tag{4.33}$$

where δ is the time interval between successive time slices. The evolution equations arise from varying $S_{\rm dis}$ with respect to h_n^m and ψ_n^m separately. For convenience, introduce the shorthand notations $h \equiv h_n^m$, and also

$$\begin{aligned} h_L &\equiv h_{n-1}^m , \qquad & h_U &\equiv h_n^{m+1} , \\ h_R &\equiv h_{n+1}^m , \qquad & h_D &\equiv h_n^{m-1} , \end{aligned} \tag{4.34}$$

and similarly for ψ. Think of L, R, U, D as meaning left, right, up and down. Then varying $S_{\rm dis}$ with respect to ψ_n^m gives

$$\psi_U = \psi - \delta^2(\psi - \psi_R) - \delta^2 \Big(\frac{n-1}{n} \Big) \Big(\frac{1 - h_L^2}{1 - h^2} \Big) (\psi - \psi_L) + \Big(\frac{1 - h_D^2}{1 - h^2} \Big) (\psi - \psi_D) , \tag{4.35}$$

while varying with respect to h_n^m gives

$$(h - h_U)(h h_U - 1) = A(1 - h^2)^2 , \tag{4.36}$$

where

$$A = \begin{cases} \delta^2 \Big(\dfrac{(h - h_R)(h h_R - 1)}{(1 - h^2)^2} + N(Nh + 1) \Big) + \Big(\dfrac{h - h_D}{1 - h_D^2} \Big) \\ \qquad + h \big(\delta^2 (\psi - \psi_R)^2 - (\psi - \psi_U)^2 \big) & (n = 1), \\ \delta^2 \Big(\dfrac{(h - h_R)(h h_R - 1)}{(1 - h^2)^2} + \dfrac{N^2 h}{n^2} - \dfrac{n-1}{n} \Big(\dfrac{h - h_L}{1 - h_L^2} \Big) \Big) + \Big(\dfrac{h - h_D}{1 - h_D^2} \Big) \\ \qquad + h \big(\delta^2 (\psi - \psi_R)^2 - (\psi - \psi_U)^2 \big) & (n > 1). \end{cases}$$

Equations (4.36) and (4.35) together form the evolution scheme for h and ψ. Note that (4.35) gives ψ_U explicitly in terms of the other quantities; but (4.36) is quadratic in h_U

(unless $h = 0$, in which case it is linear). The correct choice for h_U comes from requiring that as $\delta \to 0$ (and hence $A \to 0$) then $h_U \to h$:

$$h_U = \begin{cases} A & (h = 0)\,, \\ \frac{1}{2h}\left(1 + h^2 - (1 - h^2)\sqrt{1 - 4hA}\right) & (h \neq 0)\,. \end{cases} \tag{4.37}$$

It should be emphasized that if one sets $h = h_U = h_D$ (*i.e.* looks for time-independent solutions on the lattice) then the Bogomolny relations (4.19) imply (4.36). It is this feature that distinguishes the scheme from a more naïve approach; in other words, one gets exact static solutions on the lattice.

To be completely rigorous, one should carry out some sort of analysis of numerical stability for the discrete model. However, (4.35) and (4.36) are sufficiently complicated to make this far from a trivial task. For the moment, one must be content with the fact that extensive use of these difference equations has not revealed any instabilities.

4.5 Boundary Conditions and Numerical Output

The lattice formulation necessarily has a spatial boundary, at $n = n_{\max}$ say. As in any numerical evolution, the boundary conditions must be chosen with care. In this case they are chosen as follows.

$$\text{For } N = 1 \qquad \begin{cases} h^{m+1}_{n_{\max}+1} = h^{m}_{n_{\max}+1}\,, \\ \psi^{m+1}_{n_{\max}+1} = \psi^{m}_{n_{\max}+1}\,. \end{cases}$$

$$\text{For } N > 1 \qquad \begin{cases} h^{m+1}_{n_{\max}+1} = h^{m+1}_{n_{\max}} - \frac{N}{n_{\max}}\left(1 - (h^{m+1}_{n_{\max}})^2\right)\,, \\ \psi^{m+1}_{n_{\max}+1} = \psi^{m+1}_{n_{\max}}\,. \end{cases}$$

In other words, for $N > 1$ the fields on the boundary behave like instanton fields, but for $N = 1$ they are fixed in time. The boundary conditions for $N = 1$ may appear somewhat severe, but they seem to be the only sensible choice for the following reasons.

There is a fundamental difference between the cases $N = 1$ and $N > 1$, which becomes apparent when one considers the so-called "slow-motion" approximation.[46,47] Briefly, this approximation consists of letting λ depend on t but not on r in equation (4.3), thereby truncating the full dynamics down to a two-dimensional manifold, parametrized by the

complex parameter λ. For $N > 1$ this idea has proved to be very useful (for example, a slightly more sophisticated version lies at the heart of the investigation of scattering processes presented in chapter VI); but when $N = 1$ the expressions for the kinetic energy are divergent and so the method cannot be used. The same feature also occurs in numerical evolutions in the sense that, if one attempts to apply boundary conditions in the $N = 1$ sector that allow the fields to change with time at arbitrarily large distances, then the total energy of the system grows rapidly and without bound. In short, one is forced to impose a fixed boundary condition.

Turning now to the question of numerical output, it will sometimes be useful to look at a full plot of the energy density; but for the most part it is sufficient to use just three global measures of the evolution, namely the total kinetic energy T, the total potential energy V, and a measure W of the configuration width. All these quantities are calculated inside the region $n = n_c$ on the spatial lattice, where n_c is a parameter to be specified in the model input. Clearly n_c cannot be chosen larger than n_{max}. In fact, in many situations it is useful to take $n_c < n_{max}$, so removing the outer part of the mesh from the calculation.

In the continuum model one has

$$\begin{aligned} T(n_c,t) &\equiv 2\pi \int_0^{n_c} r\epsilon_T(r,t)\,dr\,, \\ V(n_c,t) &\equiv 2\pi \int_0^{n_c} r\epsilon_V(r,t)\,dr\,, \\ W(n_c,t) &\equiv \frac{2\pi \int_0^{n_c} r^2\big(\epsilon_T(r,t)+\epsilon_V(r,t)\big)\,dr}{T(n_c,t)+V(n_c,t)}\,, \end{aligned} \tag{4.38}$$

where $\epsilon_T(r,t)$ and $\epsilon_V(r,t)$ are given by equation (4.30). The discrete versions of T, V, W come from the discrete action (4.33):

$$\begin{aligned} T_{dis}(n_c,m) &\equiv 2\pi \sum_{n=1}^{n_c} n\epsilon_T(n,m)\,, \\ V_{dis}(n_c,m) &\equiv 2\pi \sum_{n=1}^{n_c} n\epsilon_V(n,m) + \pi N(1-h_1^m)\,, \\ W_{dis}(n_c,m) &\equiv \frac{2\pi \sum_{n=1}^{n_c} n^2\big(\epsilon_T(n,m)+\epsilon_V(n,m)\big)}{T_{dis}(n_c,m)+V_{dis}(n_c,m)}\,, \end{aligned} \tag{4.39}$$

where

$$\epsilon_T(n,m) \equiv \frac{1}{4\delta^2}\left(\frac{(h_n^{m+1}-h_n^m)^2}{1-h_n^{m2}} + (1-h_n^{m2})(\psi_n^{m+1}-\psi_n^m)^2\right),$$
$$\epsilon_V(n,m) \equiv \frac{1}{4}\left(\frac{(h_{n+1}^m-h_n^m)^2}{1-h_n^{m2}} + \frac{N^2(1-h_n^{m2})}{n^2} + (1-h_n^{m2})(\psi_{n+1}^m-\psi_n^m)^2\right). \tag{4.40}$$

4.6 Preliminary Results

This section checks the numerical evolution in the simple cases of a charge-one instanton and a slowly moving charge-two ring. Having already decided on boundary conditions in the previous section, the main discussion here centres on the choice of initial data. We will find that using the discrete Bogomolny relations to construct initial data is much better then taking values directly from the continuum solutions.

Figure 4.1 deals with the charge-one instanton given by $\lambda = 30$ evolved over the range $0 \le t \le 100$, and with $n_c = n_{max} = 100$ and $\delta = 0.1$. Recall that the continuum solution is $\psi(r,t) = 0$ together with

$$h(r,t) = \frac{\lambda^2 - r^2}{\lambda^2 + r^2},$$
$$R(r,t) = \frac{2\lambda r}{\lambda^2 + r^2}, \tag{4.41}$$

which, when substituted into (4.38), leads to $T(n_c,t) = 0$ and

$$V(n_c,t) = \frac{2\pi n_c^2}{\lambda^2 + n_c^2},$$
$$W(n_c,t) = \frac{\lambda}{n_c^2}\left((\lambda^2 + n_c^2)\tan^{-1}\frac{n_c}{\lambda} - \lambda n_c\right). \tag{4.42}$$

These last two quantities are plotted in figure 4.1 as finely broken lines.

The other curves are the results of numerical evolutions, for two different sets of initial data. Since the evolution equations are second order, the initial data must specify the field values at the first two time slices, $m = 0$ and $m = 1$ say. Suppose that $h_n(\lambda)$ is the

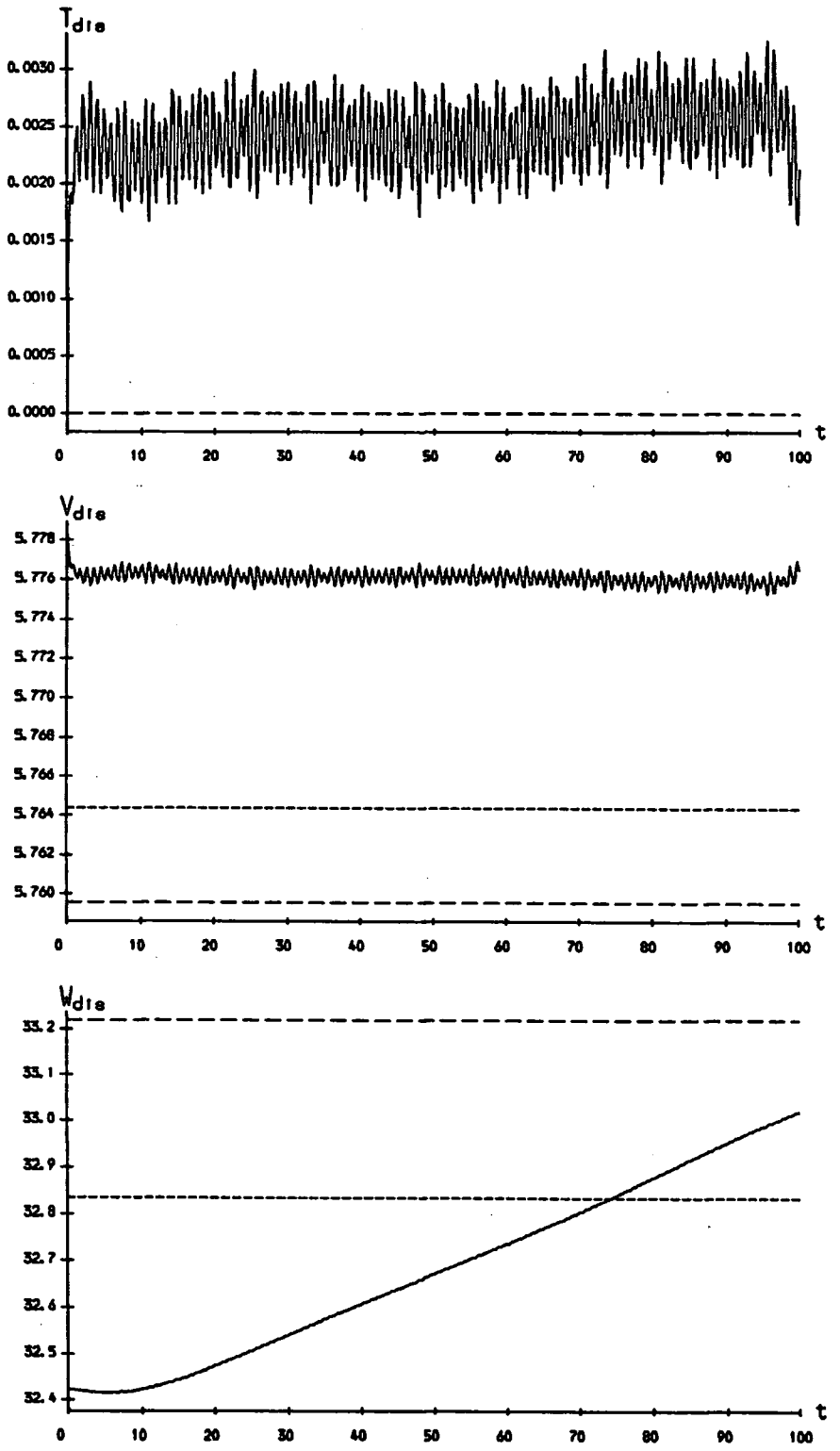


FIG 4.1. The variations of T, V and W over the range $0 \leq t \leq 100$ for a charge-one instanton in the continuum model (finely broken line), and also for the numerical evolution. In the latter case, initial data are taken both directly from the continuum model (solid line) and from the discrete Bogomolny relations (coarsely broken line).

value of h at lattice site n for a lattice instanton of width λ. The coarsely broken line is derived from initial data given by the discrete Bogomolny relations:

$$\begin{aligned} h_n^0 &= h_n^1 = h_n(\lambda)\,, \\ \psi_n^0 &= \psi_n^1 = 0\,. \end{aligned} \tag{4.43}$$

As a comparison, the solid curve is derived from initial data taken directly from the continuum solution:

$$\begin{aligned} h_n^0 &= h_n^1 = (\lambda^2 - n^2)/(\lambda^2 + n^2)\,, \\ \psi_n^0 &= \psi_n^1 = 0\,. \end{aligned} \tag{4.44}$$

The most striking feature is that initial data taken from the continuum model lead to "lattice wobble", which is eliminated when the discrete Bogomolny relations are used instead. Of course, this is precisely what the discrete Bogomolny formalism was designed to do.

Note from figure 4.1 that the Bogomolny relations lead to a value of W which is slightly larger than that predicted by the continuum analysis (a feature caused by λ being redefined on the lattice to be the limit of $\{\lambda_n\}$). The absence of a natural scale in the problem means that this small difference is unimportant; what matters is that taking initial data from the Bogomolny relations leads to numerical results which are qualitatively close to the continuum model.

Turning now to time-dependent configurations, one could attempt to reproduce some of the solutions which have been obtained analytically for $N = 0$. (Included in these is the special case $h = 0$, with ψ satisfying the radial wave equation.) However, taking $N = 0$ does not test the ability of the model to handle nontrivial topologies. Instead we shall consider evolutions in the charge-two sector obtained using the slow-motion approximation. One such evolution[46] is given by

$$\lambda(t) = a(b + it)^2\,, \tag{4.45}$$

where a and b are real constants. The kinetic energy is $\pi^2 a$ (so the approximation is expected to be best when a is small) and the potential energy is 4π. Physically, this solution corresponds to a ring, peaked at $r = 3^{-1/4}\sqrt{a(b^2 + t^2)}$, which contracts to a minimum radius at $t = 0$ and then expands again. Note that because the evolution is being approximated by a sequence of instanton configurations, the kinetic and potential

energies are conserved separately. Figures 4.2 and 4.3 show this evolution for $a = 0.001$ and $b = 1000$, over the range $0 \leq t \leq 1000$ and taking $n_c = 150$. In figure 4.2, n_{max} is set equal to 150, while in figure 4.3 it is increased to 300, *i.e.* figure 4.3 investigates boundary effects, by putting the boundary further away but keeping n_c fixed.

In the continuum model, (4.45) corresponds to

$$h(r,t) = \frac{\gamma^4 - r^4}{\gamma^4 + r^4}\,,$$

$$R(r,t) = \frac{2\gamma^2 r^2}{\gamma^4 + r^4}\,, \tag{4.46}$$

$$\psi(r,t) = \arctan\left(\frac{2bt}{b^2 - t^2}\right)\,,$$

where $\gamma = \sqrt{a(b^2 + t^2)}$. Substituting these expressions into (4.38) yields

$$T(n_c,t) = 2a\pi\left(\frac{\pi}{2} - \tan^{-1}\frac{\gamma^2}{n_c^2} - \frac{\gamma^2 n_c^2}{\gamma^4 + n_c^4}\right)\,, \tag{4.47}$$

$$V(n_c,t) = \frac{4\pi n_c^4}{\gamma^4 + n_c^4}\,, \tag{4.48}$$

$$W(n_c,t) = \tfrac{1}{4}\sqrt{2}\gamma\left(1 + \frac{\gamma^4}{n_c^4}\right)\left(\tanh^{-1}\frac{\sqrt{2}\gamma n_c}{\gamma^2 + n_c^2} + \arctan\frac{\sqrt{2}\gamma n_c}{\gamma^2 - n_c^2}\right) - \frac{\gamma^4}{n_c^3}\,, \tag{4.49}$$

where in (4.49) the principal range of arctan is taken to be $[0,\pi)$. These quantities are plotted as finely broken lines in figures 4.2 and 4.3.

As before, the other two curves are evolved numerically from two different sets of initial data. Firstly, the coarsely broken line shows the evolution of initial data derived from the discrete Bogomolny equations:

$$h_n^0 = h_n(ab^2)\,, \qquad h_n^1 = h_n\big(a(b^2 + \delta^2)\big)\,.$$

$$\psi_n^0 = 0\,, \qquad \psi_n^1 = \arctan\left(\frac{2b\delta}{b^2 - \delta^2}\right)\,.$$

As a comparison, the solid line shows the evolution of initial data taken directly from the

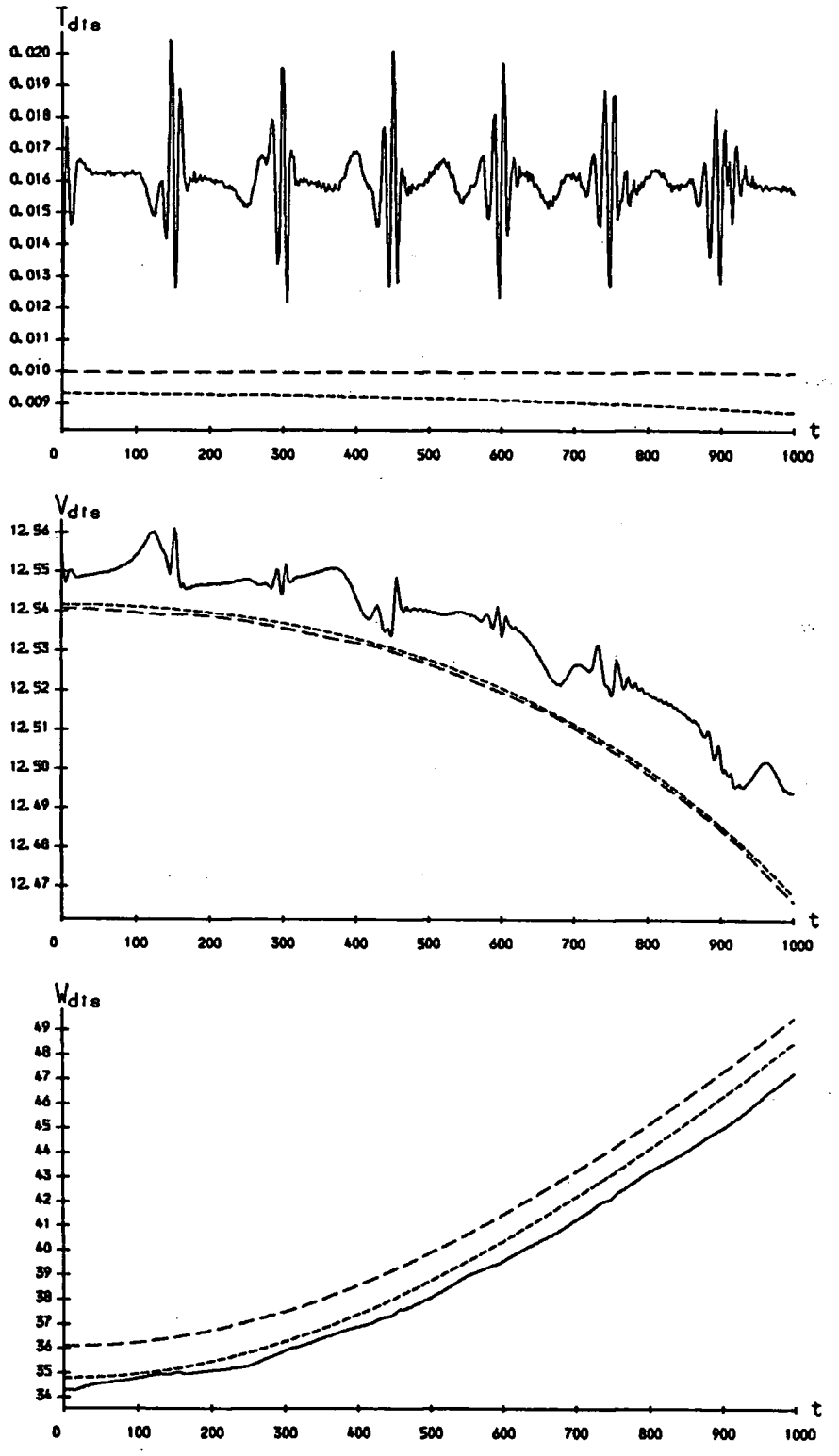


FIG 4.2. The variations of T, V and W over the range $0 \leq t \leq 1000$ for a slowly expanding charge-two ring in the analytic slow-motion approximation (finely broken line), and also for the numerical evolution. In the latter case, initial data are taken both directly from the slow-motion analysis (solid line) and from the discrete Bogomolny relations (coarsely broken line).

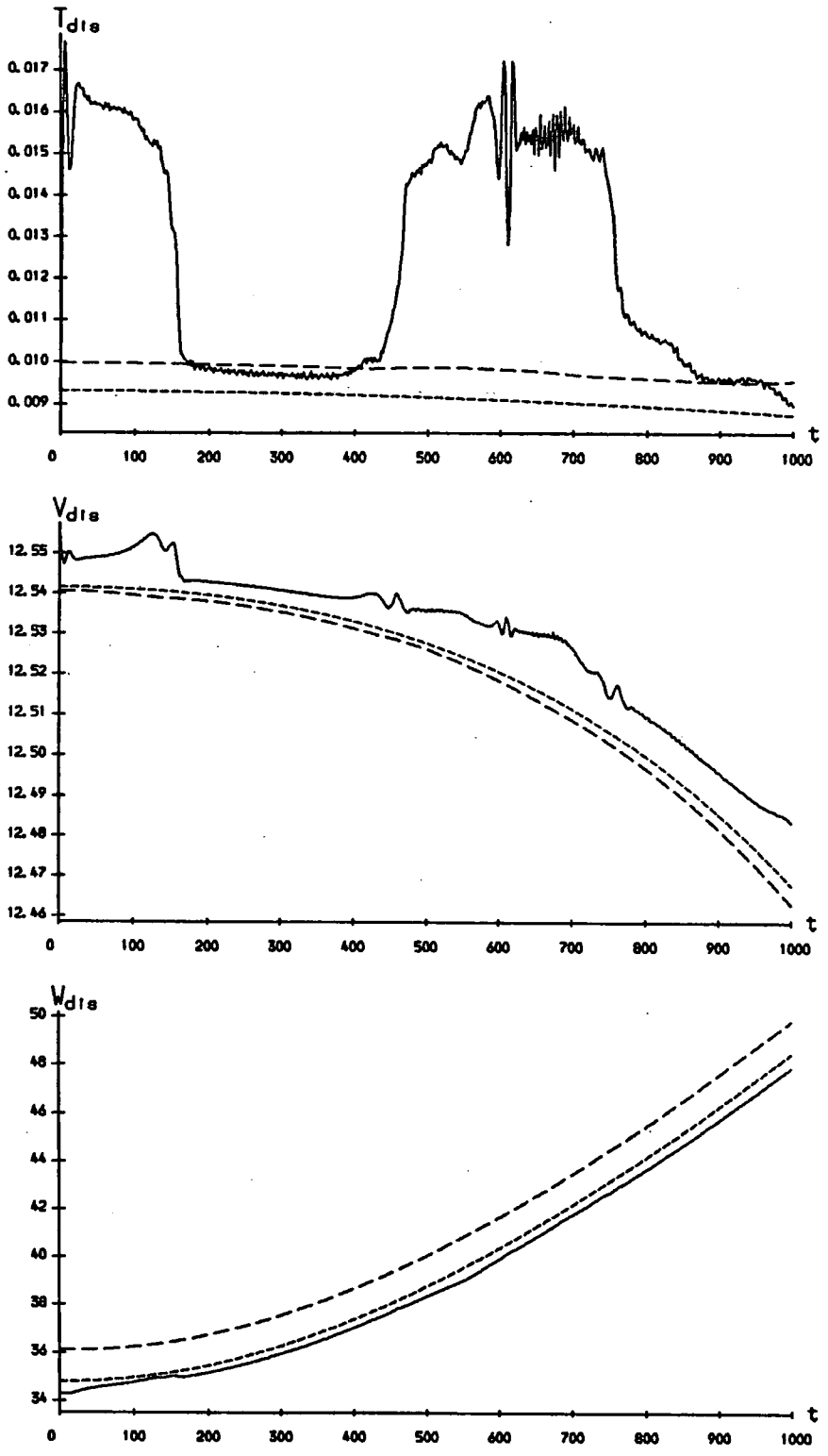


FIG 4.3. The variations of T, V and W for a slowly expanding charge-two ring. The parameters are the same as in figure 4.2, except that, in the numerical evolution, $n_{\max}$ is set to 300, instead of 150.

continuum solution:

$$h_n^0 = \frac{a^2b^4 - n^4}{a^2b^4 + n^4}, \qquad h_n^1 = \frac{a^2(b^2+\delta^2)^2 - n^4}{a^2(b^2+\delta^2)^2 + n^4}.$$

$$\psi_n^0 = 0, \qquad \psi_n^1 = \arctan\left(\frac{2b\delta}{b^2 - \delta^2}\right).$$

Once again, taking initial data directly from the continuum solution leads to lattice wobble. On the other hand, using the Bogomolny relations gives a much smoother numerical evolution.

It is interesting to note that, since the fields are not fixed at the boundary in the charge-two sector, energy may flow off the edge of the lattice (or equally into the lattice from outside). In figures 4.2 and 4.3 approximately 0.75% of the original energy is "lost" in this way as the ring expands up to $t = 1000$.

When the boundary is moved back (figure 4.3) there is very little change in the evolutions of V and W. The greatest change from figure 4.2 occurs in T: the solid line has a different shape, although it still fluctuates wildly, and the coarsely broken line now exhibits the gradual decrease predicted by the slow motion approximation. In short, it seems that to a large extent the boundary conditions are transparent to the flow of energy; but they are not completely invisible. However, since we using only a one-dimensional lattice, it is usually computationally feasible to make the boundary effectively invisible by taking $n_{\max} \gg n_c$, and we shall use this technique in the next chapter.

4.7 Concluding Remarks

The development of a set of discrete Bogomolny relations removes the lattice wobble that is observed in numerical simulations at low kinetic energies. Two particular cases have been studied in some detail and the numerical scheme appears to be working very well. In the next chapter we proceed to a full investigation of soliton stability. It may turn out that the absence of a natural scale in the model means that if a soliton is squashed then it eventually becomes an infinitely tall spike. On the other hand, the requirement of finite kinetic energy means that the fields must be fixed at infinity, and this may rescue the situation. Clearly it is essential that the lattice wobble be eliminated if one is to study small perturbations.

Finally, we remark that the idea of discretizing the Bogomolny bound, in order to obtain static solutions on the lattice, may be applied to other models that have non-trivial topologies. A few possibilities are the sine-Gordon equation in (1+1) dimensions, the Maxwell-Higgs model in (2+1) dimensions (which can describe vortices in superconductors) and the Skyrme model in (3+1) dimensions. Although there have been many numerical studies (see, for example, references [48]–[52]), this approach to a discrete evolution scheme does not seem to have been considered up till now.

Chapter V

Soliton Stability in the O(3) Sigma Model

In this chapter the question of soliton stability in the O(3) σ-model is investigated numerically, using the evolution scheme developed in the previous chapter. All the calculations deal with the evolution of a single soliton of topological charge one. Correspondingly, the boundary conditions are such that the fields are fixed in time at $n = n_{\max}$. Qualitatively at least, all the results have been confirmed by Zakrzewski and Peyrard[3] using a conventional Runge-Kutta technique.

5.1 Introduction

There are very many different types of perturbation which could be applied to an instanton solution, the only restriction being that the field is not perturbed close to the boundary. As before, suppose that $h_n(\lambda)$ is the value of h at lattice site n for a lattice instanton of width λ. Throughout this chapter the field configuration at time slice $m = 0$ is always taken to be a lattice instanton, *i.e.*

$$h_n^0 = h_n(\lambda)\,, \quad \psi_n^0 = 0\,, \tag{5.1}$$

and the perturbation is encoded in the time slice $m = 1$. We shall consider perturbations specified by four parameters, v_h, v_ψ, n_0 and n_1 (where $n_0 \leq n_1$), in the following way:

$$h_n^1 \equiv \begin{cases} h_n(\lambda + \delta v_\lambda) & (n \leq n_0) \\ h_n(\lambda + \delta v_\lambda)\left(\dfrac{n_1 - n}{n_1 - n_0}\right) + h_n(\lambda)\left(\dfrac{n - n_0}{n_1 - n_0}\right) & (n_0 < n < n_1) \\ h_n(\lambda) & (n \geq n_1) \end{cases}$$

$$\psi_n^1 \equiv \begin{cases} \delta v_\psi & (n \leq n_0) \\ \delta v_\psi\left(\dfrac{n_1 - n}{n_1 - n_0}\right) & (n_0 < n < n_1) \\ 0 & (n \geq n_1)\,. \end{cases} \tag{5.2}$$

Physically the picture is this. Inside $n = n_0$, v_λ is the initial rate of change of the soliton width; $v_\lambda < 0$ corresponds to an initial contraction and $v_\lambda > 0$ to an initial expansion. Similarly, v_ψ is the initial time derivative of ψ. Outside $n = n_1$ there is no perturbation at all. In the region $n_0 < n < n_1$ there is a continuous interpolation between the inner and outer regions. It is expected that this class of perturbations is sufficiently large to reveal all the qualitative types of behaviour that can occur.

Before discussing any results in detail, recall that all the numerical output is calculated from the field values inside a radius $n = n_c$. It is therefore useful to truncate the lattice analogue of the Bogomolny bound down to the same radius. In the notation of the previous chapter, it is easy to show that

$$V_{\rm dis}(n_c, m) \geq \pi N(1 - h_{n_c+1}), \tag{5.3}$$

with equality if and only if the field configuration is a lattice instanton. Note that provided $n_c \neq n_{\rm max}$, the truncated Bogomolny bound may change as the fields evolve in time. In many cases much can be learned by plotting this bound on the same axes as the potential energy.

5.2 Numerical Results

We shall discuss four numerical evolutions in some detail. The first one takes place on a relatively small mesh and illustrates the effects that can arise from the boundary. The other three address the main question of stability on an infinite plane, and each reveals a different type of behaviour.

The first evolution takes place on a mesh of size $n_{\rm max} = 200$, on which is placed a soliton of width 30. The initial perturbation is of the form (5.2) with

$$n_0 = 30, \quad n_1 = 60, \quad v_\lambda = -0.5, \quad v_\psi = 0.0. \tag{5.4}$$

The kinetic energy is initially about 5% of the total. The configuration was evolved up to $t = 4000$ with a time step $\delta = 0.5$. Using double precision arithmetic on the Amdahl mainframe in Durham, this took about 12 minutes of CPU time. The total energy was conserved to within 0.01%. Note that, since $v_\psi = 0$, ψ is identically zero throughout the

calculation and the whole evolution is accounted for by the variation of h. Figure 5.1 shows the kinetic energy density during the early stages. The first picture is the initial perturbation, which takes the form of a hollow ring (the spikes around the rim are due to the graphics package); subsequent pictures show a large burst of radiation travelling outwards at the speed of light, together with a residual motion in the central region occupied by the soliton. In the last picture the radiation has just reached the boundary at $n = 200$. After this it is reflected back, reabsorbed by the soliton and then another pulse is emitted a short time later; and so the process repeats. Figure 5.2 shows the time variation of $W_{\rm dis}$, taking $n_{\rm c} = 200$. There is a clear periodic behaviour, caused by the burst of radiation being repeatedly reflected back in from the boundary.

We turn now to stability on the infinite plane. In order to eliminate boundary effects, we shall use a technique that is effective rather than elegant: simply take $n_{\rm max} \gg n_{\rm c}$. Then, provided we do not evolve beyond $t = 2(n_{\rm max} - n_{\rm c})$, there is no time for radiation to travel out to the boundary and be reflected back inside the region $n < n_{\rm c}$. Of course, this is computationally expensive, but, because axial symmetry reduces the theory to (1+1) dimensions, it is perfectly feasible, and the advantages are twofold. Firstly, one can be sure that boundary effects do not come into play and second, it creates the illusion of radiation "falling off" the edge of the mesh. Figure 5.3 presents results for $n_{\rm max} = 6000$, $n_{\rm c} = 1000$ and $\delta = 0.5$, with the initial soliton having $\lambda = 50$ and perturbation given by

$$n_0 = 100, \quad n_1 = 175, \quad v_\lambda = -0.05, \quad v_\psi = 0.0\,. \tag{5.5}$$

The initial sharp drop in kinetic energy corresponds to the emission of a pulse of radiation, and the step around $t = 1000$ is caused by the radiation spreading outside $n = n_{\rm c}$. In rough terms, the emission of radiation causes a drop of about 40% in the kinetic energy; the pulse itself carries away another 50%, leaving a residue of about 10% in the region occupied by the soliton. The graph of $W_{\rm dis}$ shows that the soliton width decreases almost linearly. The numerics break down at $t = 7380$, by which time $W_{\rm dis} = 1.35$, *i.e.* the soliton occupies essentially only one lattice site. Further computer experiments show that if the perturbation is more localised then a greater proportion of the initial kinetic energy is carried away by radiation. But there is only ever one burst of radiation and there is always some residual kinetic energy, how ever small, which will eventually cause the soliton to shrink to a spike (or to expand without limit in the case $v_\lambda > 0$).

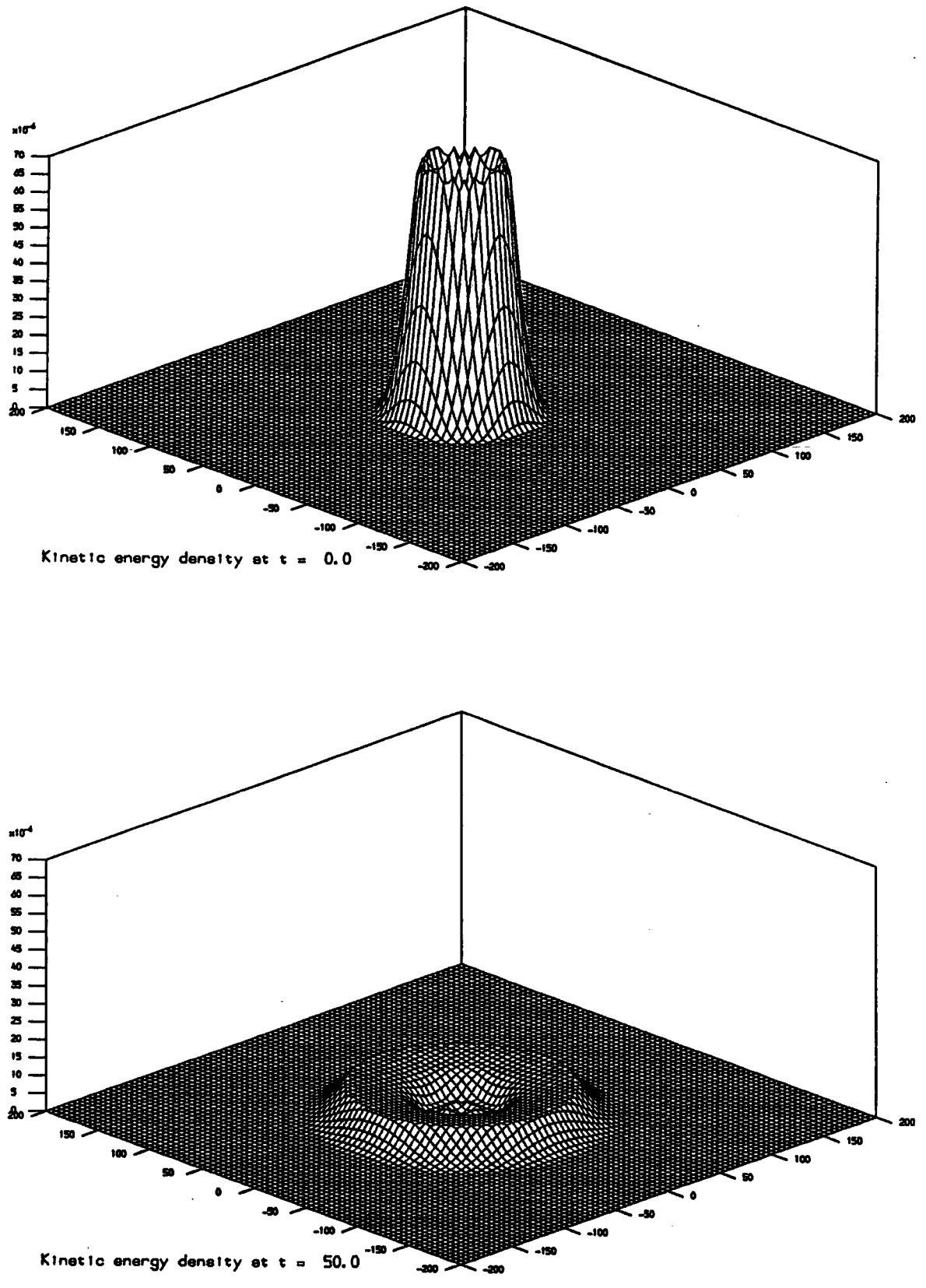
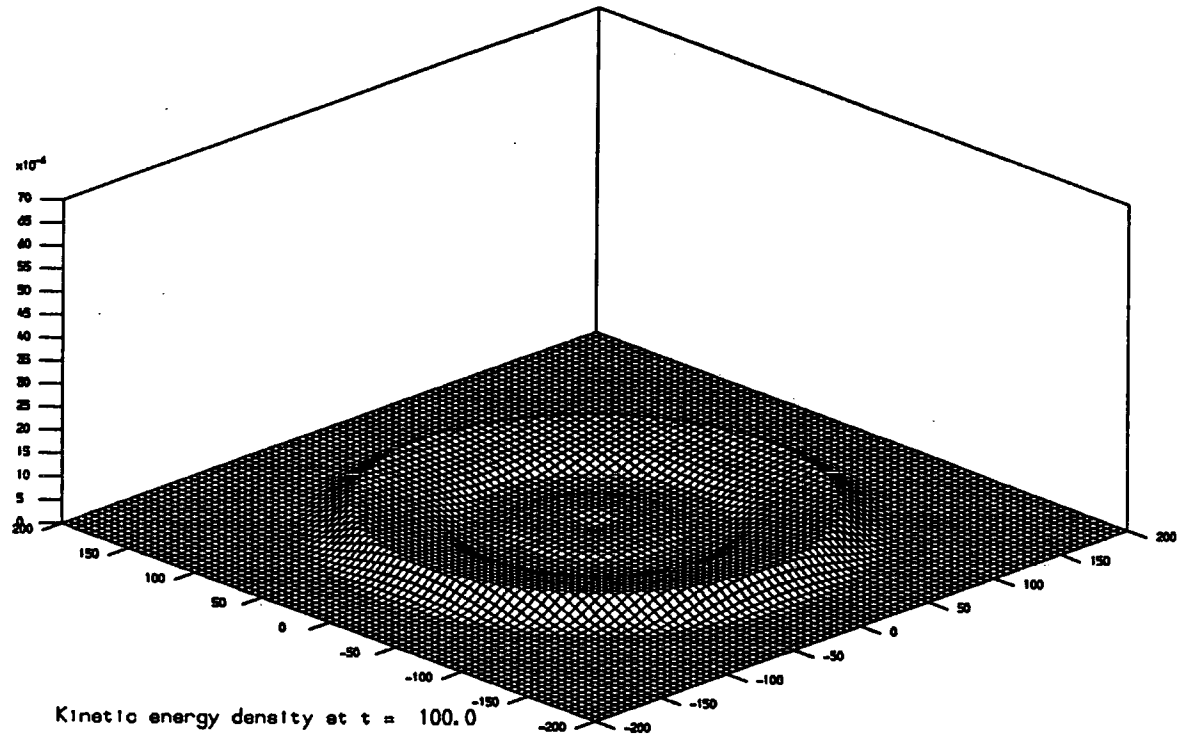


FIG 5.1. Pictures of the kinetic energy density obtained by evolving the initial perturbation (5.4) on the finite mesh $n_{\max} = 200$ with an initial soliton of width 30.



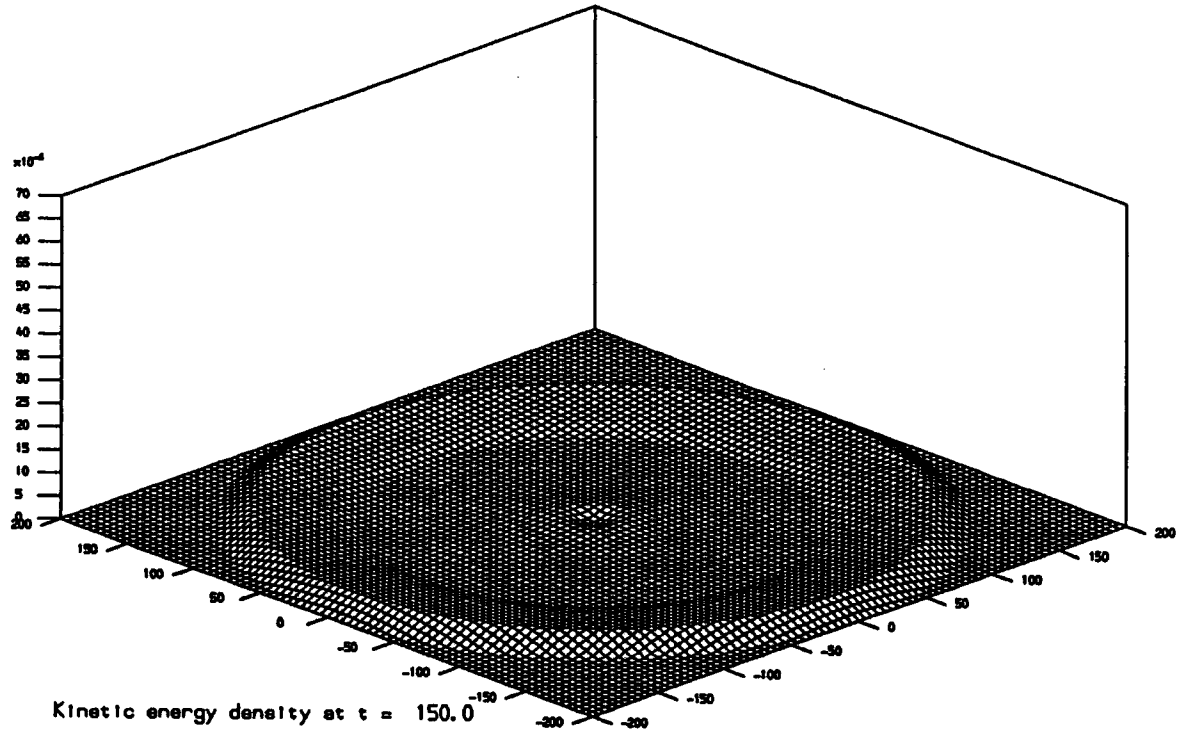


FIG 5.1. Continued.

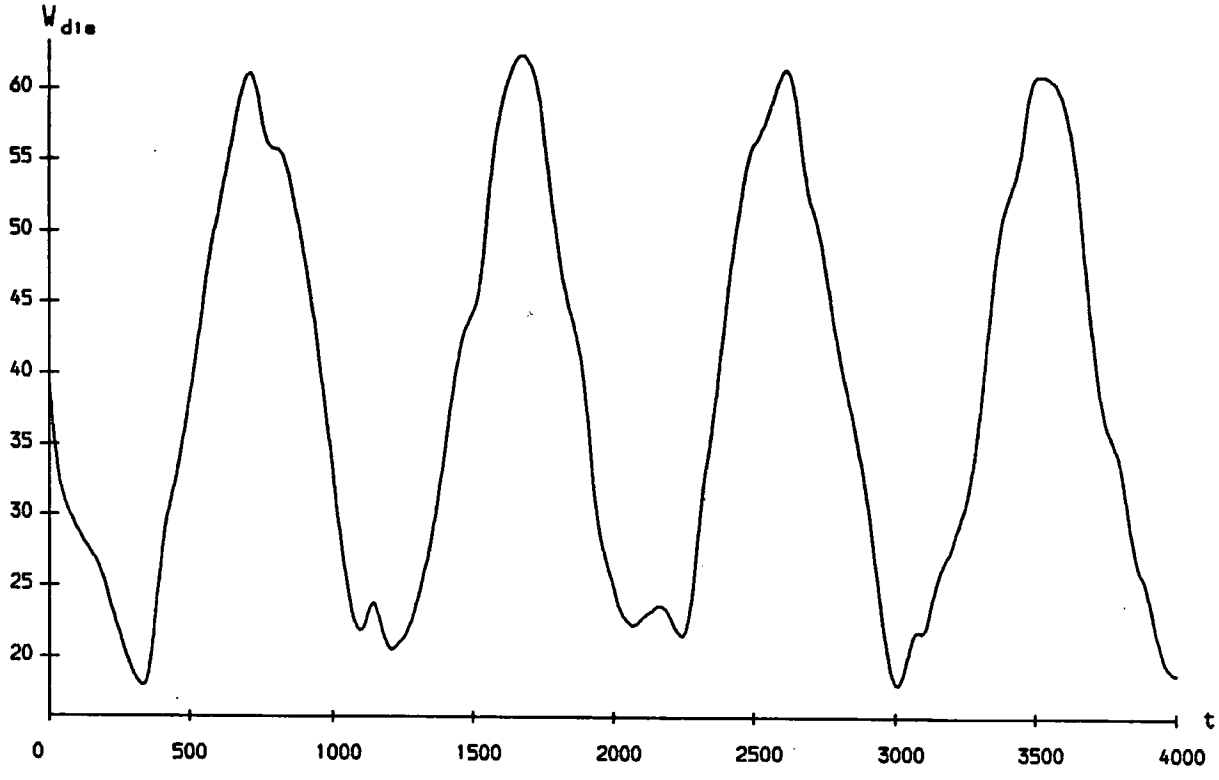


FIG 5.2. The time dependence of $W_{\rm dis}$ for the soliton of figure 5.1, taking $n_{\rm c} = 200$.

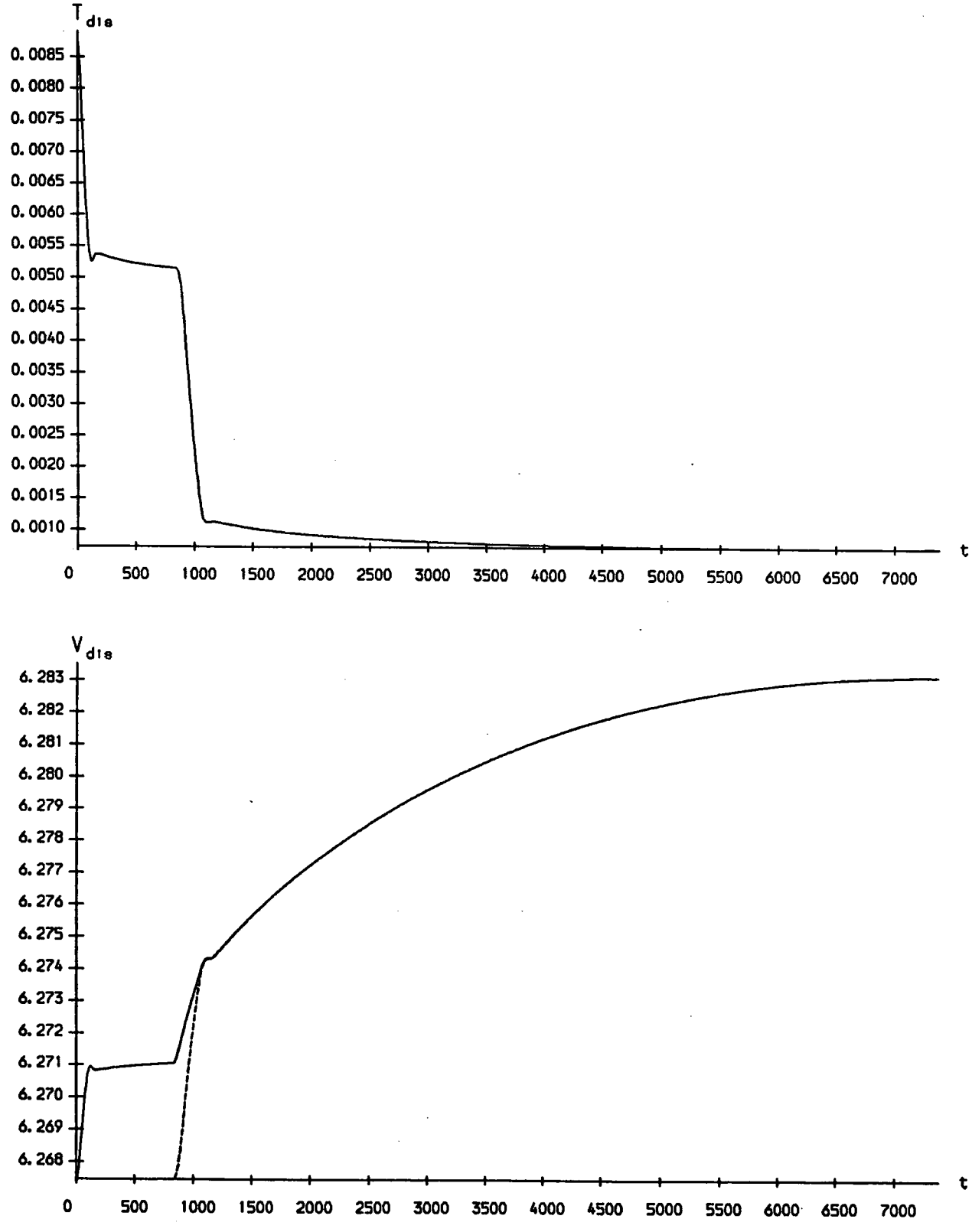


FIG 5.3. The evolutions of T_{dis}, V_{dis} and W_{dis} on the "infinite" mesh $n_{\max} = 6000$ for an initial soliton of width 50 and perturbation (5.5), and taking $n_c = 1000$. The discrete Bogomolny bound is shown as a broken line on the same axes as the potential energy.

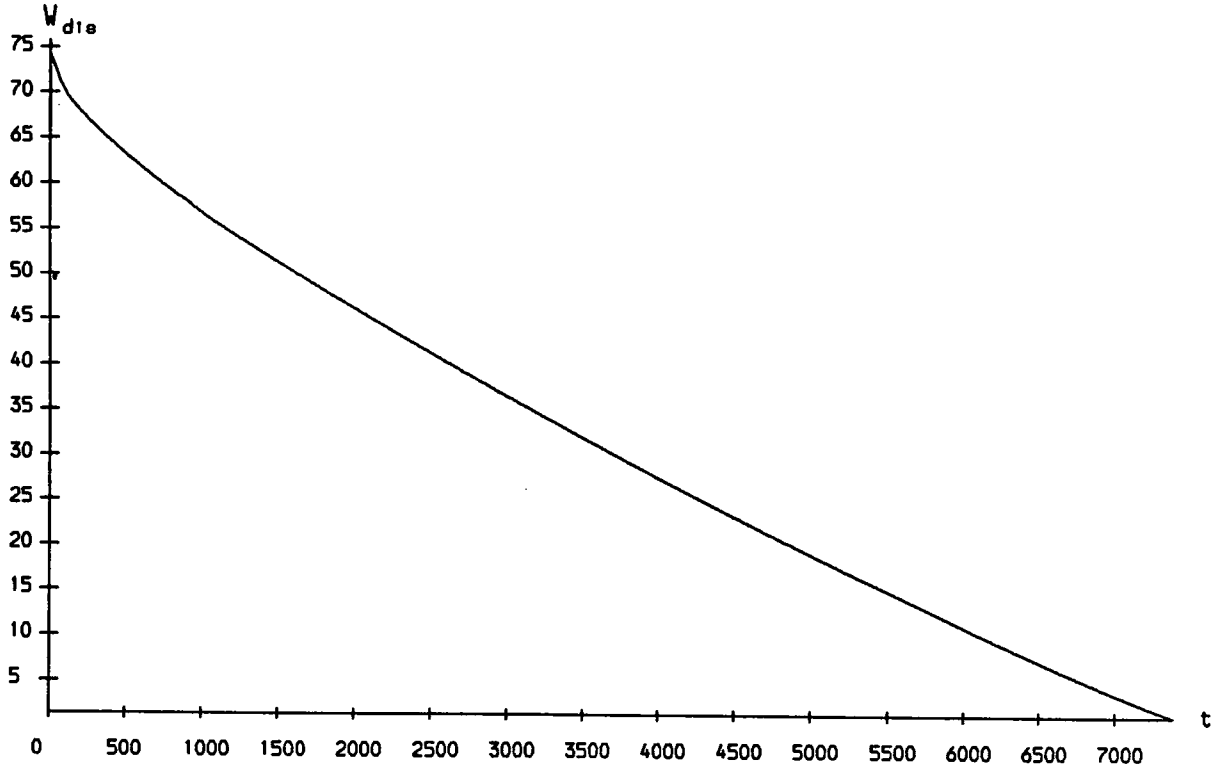


FIG 5.3. Continued.

The graph of the potential energy gives an added insight into the evolution, particularly when plotted alongside the truncated Bogomolny bound (5.3). Beyond $t = 1000$, when the radiation has been removed from the calculation, the field configuration saturates the Bogomolny bound very closely. Moreover, as the soliton shrinks, the amount of potential energy lying inside $n = n_c$ increases, eventually tending to 2π as the soliton becomes more spiky.

We move on now to the next set of results. So far, both perturbations have had $v_\psi = 0$ in (5.2) and consequently $\psi = 0$ throughout the evolution. Figure 5.4 presents results for the same mesh and initial soliton as figure 5.3, but now with a perturbation given by

$$n_0 = 100, \quad n_1 = 175, \quad v_\lambda = 0, \quad v_\psi = 0.0005\,. \tag{5.6}$$

The kinetic energy exhibits a very similar behaviour to the previous case: again a single burst of radiation is emitted, which is removed from the calculation at about $t = 1000$, after which point the Bogomolny bound is closely saturated. The essential difference is that the soliton now slowly expands, and appears to do so indefinitely.

To understand this evolution further, it is useful to split the kinetic energy into two pieces, corresponding to looking at h_t and ψ_t separately:

$$T = T_h + T_\psi\,, \tag{5.7}$$

where

$$\begin{aligned} T_h &\equiv \frac{1}{4}\int_0^\infty \frac{h_t^2}{1-h^2}\,(2\pi r)\,dr \\ T_\psi &\equiv \frac{1}{4}\int_0^\infty \psi_t^2\,(1-h^2)\,(2\pi r)\,dr\,. \end{aligned} \tag{5.8}$$

We have seen that if T_ψ is initially zero then it remains zero. On the other hand, T_h is in general not zero, even if it vanishes initially, as is the case here. So, for perturbations in which $v_\psi \neq 0$, it is useful to consider the way in which T_h/T and T_ψ/T vary with time. These quantities are plotted in figure 5.4. It is found that T_ψ/T gradually decreases (and correspondingly T_h/T increases). For example, at $t = 10000$ the time variation of ψ accounts for only about 75% of the total kinetic energy.

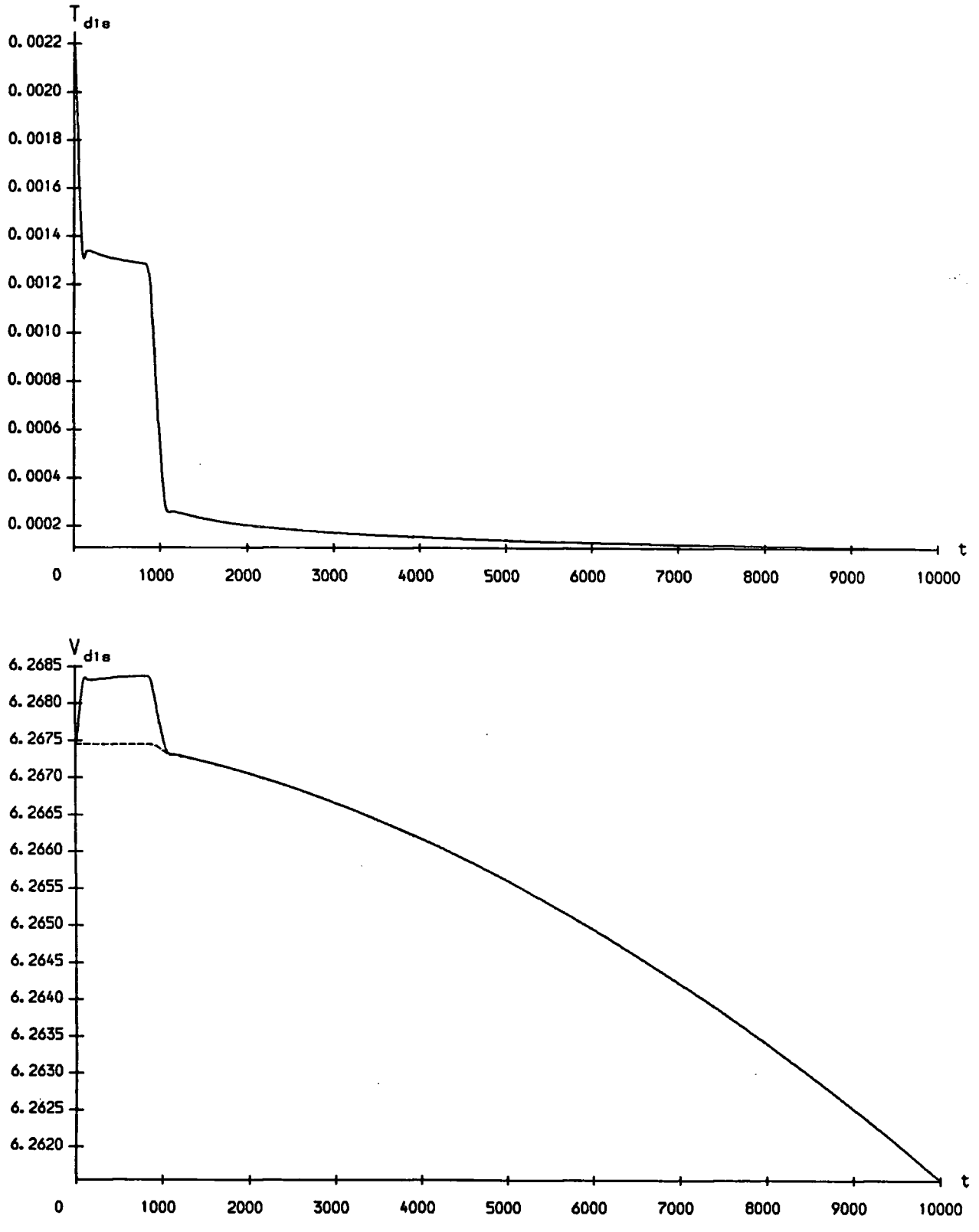


FIG 5.4. The evolution of the initial perturbation (5.6), using the same mesh as in figure 5.3 and now including a graph of the normalised kinetic energy, which shows the variation with time of $100\,T_\psi/T$ (solid line) and of $100\,T_h/T$ (broken line).

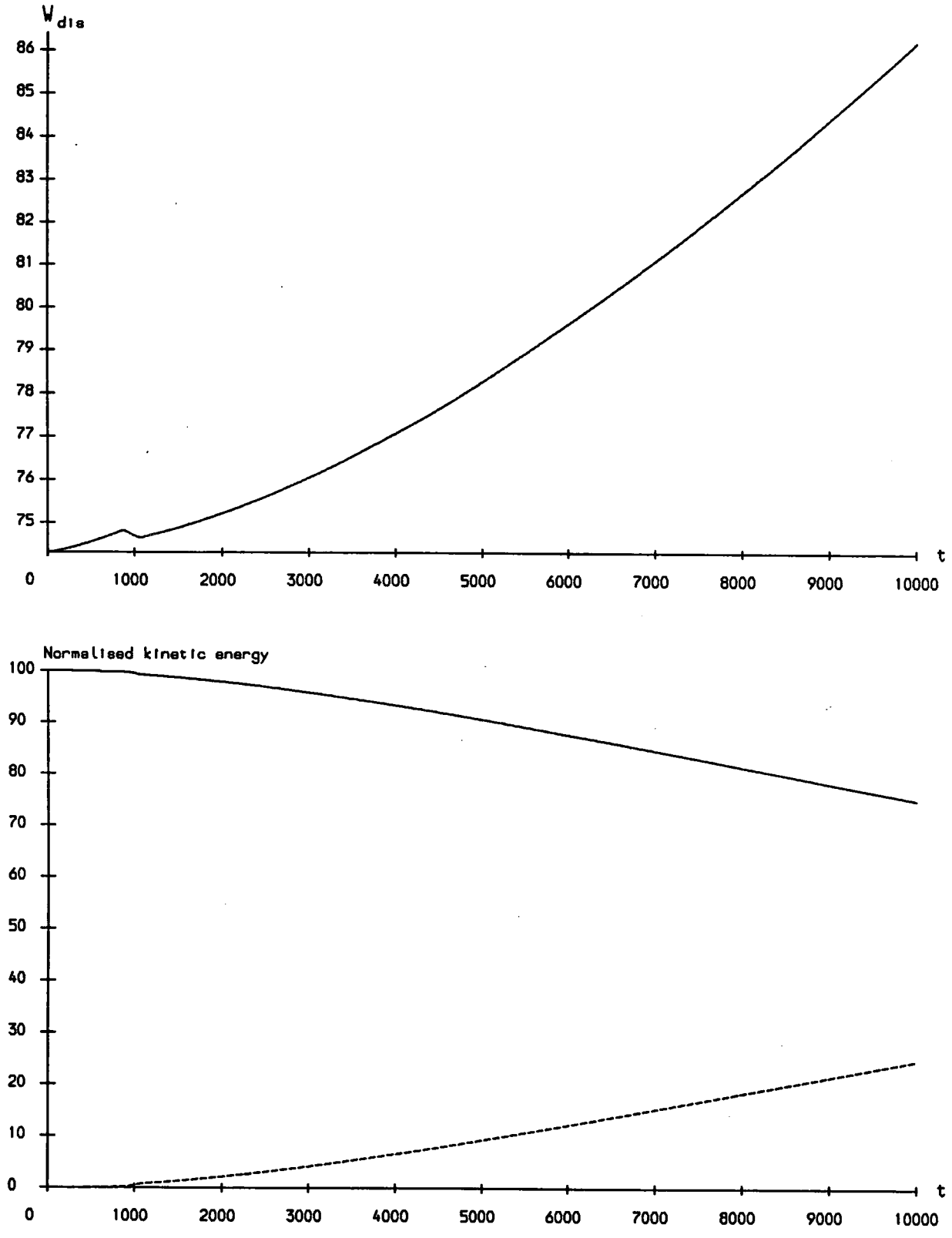


FIG 5.4. Continued.

The last type of evolution that we look at has both v_λ and v_ψ nonzero. The case when $v_\lambda < 0$ is of particular interest. We have seen how the evolutions due to v_λ and v_ψ separately cause the soliton to shrink and expand respectively, but it is not clear at first sight which effect will dominate when both are present. Figure 5.5 shows results for the same mesh and initial soliton as figures 5.3 and 5.4, but now with a perturbation given by

$$n_0 = 100, \quad n_1 = 175, \quad v_\lambda = -0.1, \quad v_\psi = 0.0005\,. \tag{5.9}$$

The total kinetic energy exhibits the now familiar fall-off, indicating the emission of radiation. The potential energy again closely saturates the Bogomolny bound once the radiation has been removed. The graph of $W_{\rm dis}$ shows that the soliton reaches a minimum width at about $t = 2800$ and then expands, apparently indefinitely. Looking at T_h/T and T_ψ/T gives some clue to the mechanism which causes this behaviour. Initially, T_ψ only accounts for about 6% of the total kinetic energy, but as the soliton shrinks this fraction increases until, at the minimum width, $T_\psi = T$. Beyond this point, T_ψ/T decreases as the soliton expands.

To sum up, the soliton overcomes its tendency to become a spike by transferring its kinetic energy to the phase ψ; but this does not stabilise it, since it eventually expands without limit. Numerical experiments with different values of v_λ and v_ψ always exhibit a similar behaviour, provided that the soliton does not shrink so much that the numerics break down before it reaches its minimum width.

As a final remark, we make the general observation that, when $v_\psi \neq 0$, a small soliton width is associated with a high value of T_ψ/T and vice-versa. This feature is a consequence of the existence of the conserved integral of motion, M, defined by

$$M \equiv \int_0^\infty \psi_t\,(1 - h^2)\,(2\pi r)\,dr\,, \tag{5.10}$$

which arises from the invariance of the action under translations of $\psi(r,t)$ by a constant. Physically, M may be thought of as the angular momentum of the soliton.

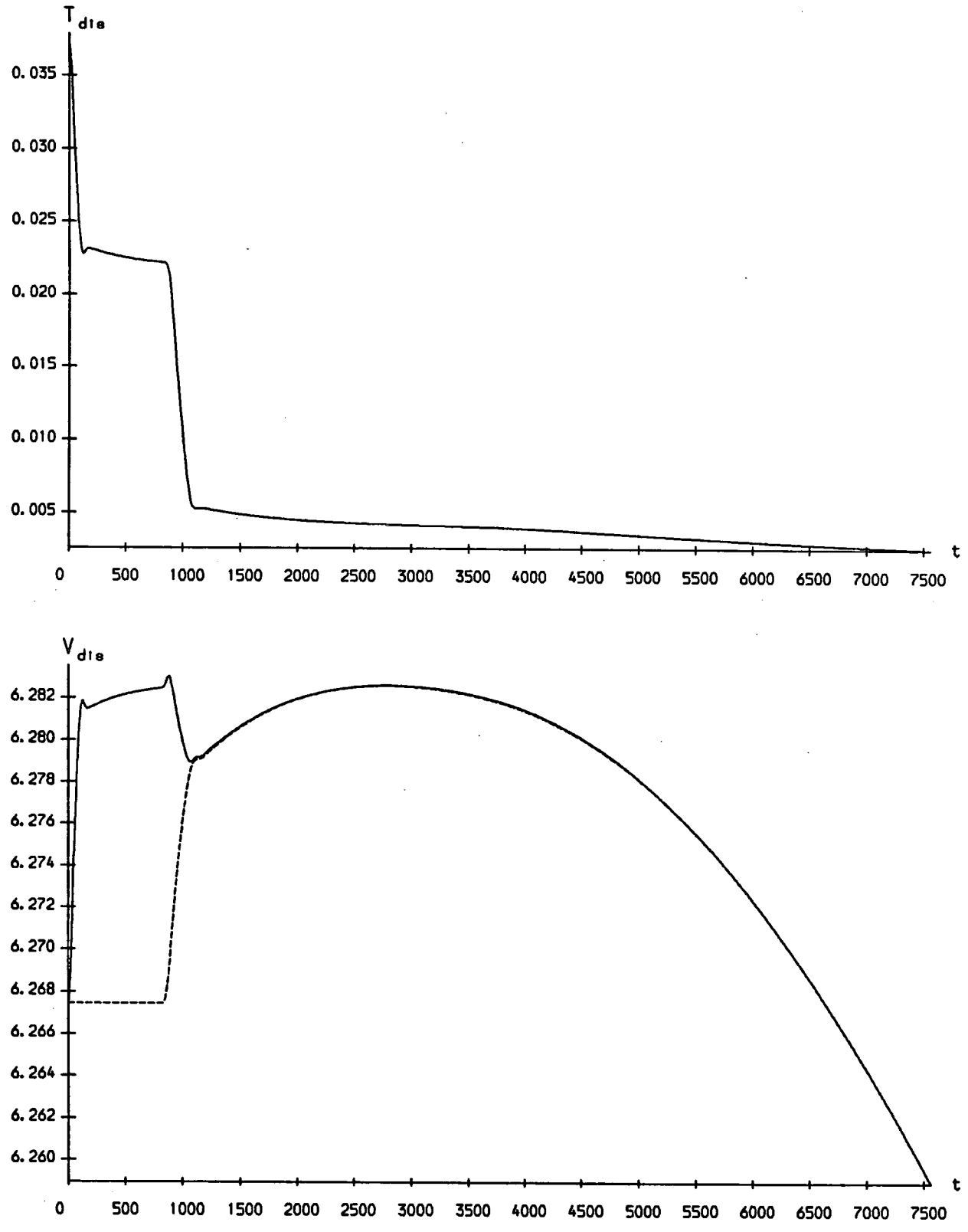


FIG 5.5. The evolution of the initial perturbation (5.9), using the same mesh as in figures 5.3 and 5.4.

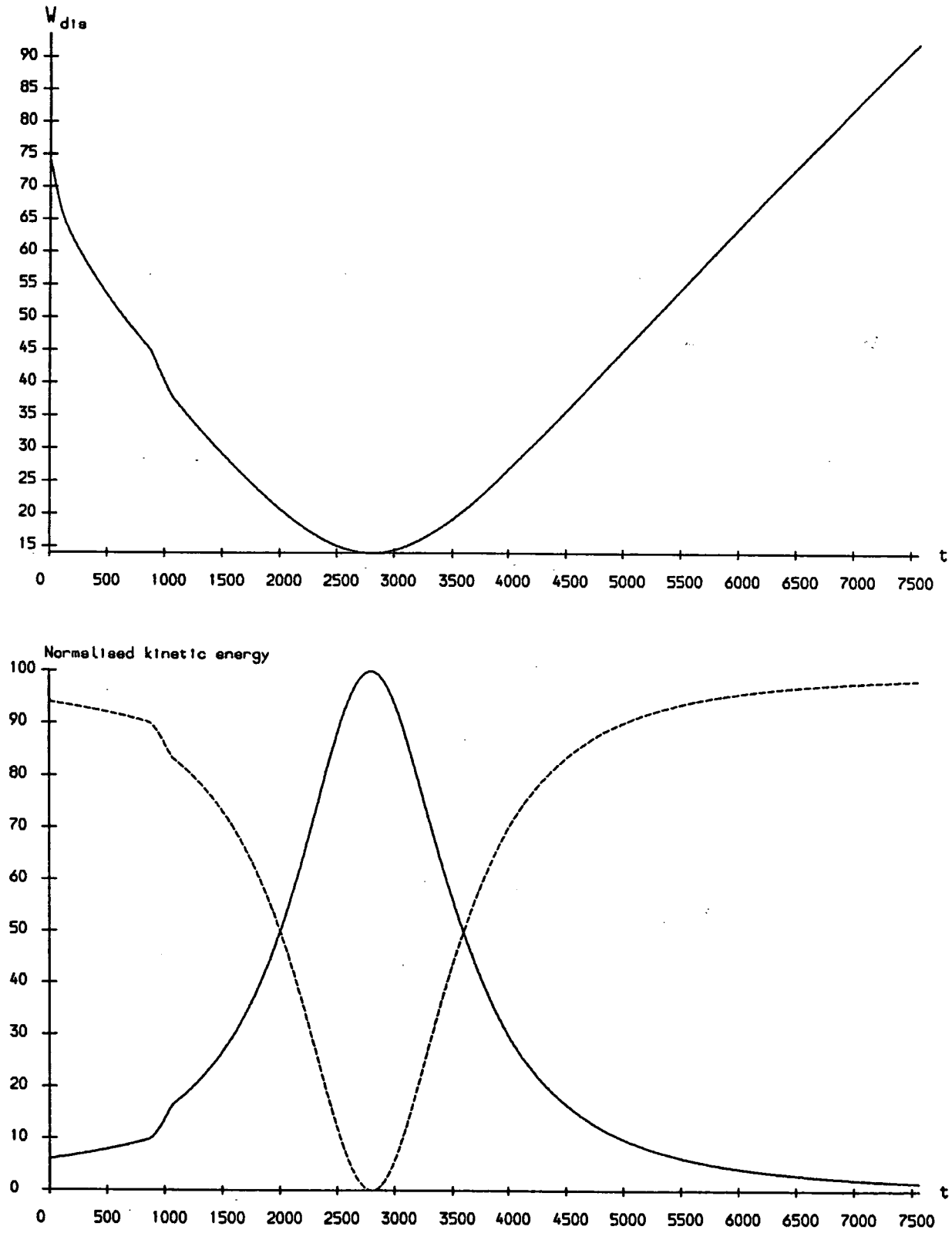


FIG 5.5. Continued.

5.3 Comparison with the Wave Equation

The main conclusion of this numerical investigation is that a very small perturbation can cause an initially smooth field configuration to become very spiky. But this is not a phenomenon that is familiar from other (2+1)-dimensional field theories; in particular it does not happen in the (axially symmetric) wave equation,

$$\phi_{rr} + \frac{\phi_r}{r} - \phi_{tt} = 0\,, \tag{5.11}$$

where ϕ is a function of r and t. Why are these two theories different? The following "energy budget" analysis does not involve solving equations of motion, but it does at least make it plausible why the two theories should exhibit different behaviours. The idea is to use the numerical results to write down an approximate form for the continuum field configurations as the soliton shrinks, and then to calculate how much energy it costs to reach these configurations from the original one. It is useful to replace $h(r,t)$ with

$$\lambda(r,t) \equiv r\left(\frac{1+h(r,t)}{1-h(r,t)}\right)^{1/2}. \tag{5.12}$$

Then the instantons correspond to taking $\lambda(r,t)$ equal to a constant, analogous to the solutions $\phi(r,t)$ constant in the wave equation. In terms of λ, the potential energy of σ-model configurations is

$$V_\sigma = \int_0^\infty (2\pi r)\left(1+\frac{\lambda^2}{r^2}\right)^{-2}\left(\frac{\lambda_r^2}{r^2} - \frac{2\lambda\lambda_r}{r^3} + \frac{2\lambda^2}{r^4}\right) dr\,. \tag{5.13}$$

Let us estimate V_σ for the shrinking soliton. Since the numerical evolution closely saturates the truncated Bogomolny bound inside $r = 1000$ (at least once the radiation is removed), it is a good approximation to take λ equal to some constant, λ_1, in this region. At large distances, the boundary condition fixes λ to be some other constant, λ_2, which gives the size of the soliton initially. So we shall consider configurations of the form

$$\lambda(r) = \begin{cases} \lambda_1 & (r \le r_1) \\ \tilde{\lambda}(r) & (r_1 < r < r_2) \\ \lambda_2 & (r \ge r_2)\,, \end{cases} \tag{5.14}$$

where $\tilde{\lambda}(r_1) = \lambda_1$ and $\tilde{\lambda}(r_2) = \lambda_2$. Since the kinetic energy is only a small fraction of the total (less than 0.1%), it is reasonable to take $\tilde{\lambda}(r)$ so as to minimise V_σ. This is a

straightforward variational problem, which is further simplified by assuming that λ_1 and λ_2 are much smaller than r_1, so that $1 + \lambda^2/r^2 \approx 1$ when $r > r_1$. It is now easy to show that the minimum value of V_σ is

$$V_\sigma = 2\pi\left(1 + \frac{2(\lambda_2 - \lambda_1)^2}{r_2^2 - r_1^2}\right), \tag{5.15}$$

corresponding to

$$\tilde{\lambda}(r) = \left(\frac{\lambda_2 - \lambda_1}{r_2^2 - r_1^2}\right)r^2 + \frac{\lambda_1 r_2^2 - \lambda_2 r_1^2}{r_2^2 - r_1^2}. \tag{5.16}$$

Now consider the same analysis in the case of the wave equation, for which the potential energy is given by

$$V_{\text{wave}} = \int_0^\infty (2\pi r)\phi_r^2\, dr\,. \tag{5.17}$$

Following the previous calculation, one considers field configurations of the form

$$\phi(r) = \begin{cases} \phi_1 & (r \le r_1) \\ \tilde{\phi}(r) & (r_1 < r < r_2) \\ \phi_2 & (r \ge r_2), \end{cases} \tag{5.18}$$

where $\tilde{\phi}(r_1) = \phi_1$ and $\tilde{\phi}(r_2) = \phi_2$. As before, one wants to choose $\tilde{\phi}(r)$ to minimise the potential energy. It turns out that this minimum value is

$$V_{\text{wave}} = 2\pi\,\frac{(\phi_2 - \phi_1)^2}{\log r_2 - \log r_1}, \tag{5.19}$$

corresponding to

$$\tilde{\phi}(r) = \frac{\phi_1 \log r_2 - \phi_2 \log r_1}{\log r_2 - \log r_1} + \frac{\phi_2 - \phi_1}{\log r_2 - \log r_1}\log r\,. \tag{5.20}$$

Equations (5.15) and (5.19) are estimates of how much energy it costs to change the field to a value λ_1 (respectively ϕ_1) inside some finite radius r_1, starting from an initial configuration in which the field takes the value λ_2 (respectively ϕ_2) over all space. The example of a soliton shrinking, as presented above, is modelled by setting $\lambda_1 = 0$, $\lambda_2 = 50$,

$r_1 = 1000$ and $r_2 = 6000$; these values give $V_\sigma \approx 9.0 \times 10^{-4}$. As a comparison, put $\phi_1 = 0$ and $\phi_2 = 50$ in (5.19) with the same r_1 and r_2; this gives $V_{\text{wave}} \approx 8.8 \times 10^3$, *i.e.* seven orders of magnitude larger than the corresponding V_σ. Of course, this discussion does not prove that the σ-model solitons are unstable, but it does show how spiky configurations in the σ-model are much more energetically favourable than might be expected.

5.4 Concluding Remarks

This study has revealed that the instantons in the O(3) σ-model are unstable to small perturbations. If the soliton does not acquire any angular momentum as the result of the perturbation, then it may shrink to a spike. On the other hand, it is conjectured that a nonzero angular momentum, how ever small, will eventually cause the soliton to expand indefinitely.

It is natural to look for ways of stabilizing the solitons. Two possibilities are currently under consideration. The first involves adding both a Skyrme term (containing fourth derivatives) and a potential term to the action density. The former tends to make the soliton expand, while the latter makes it shrink. It is the combination of both these effects which leads to a stable configuration, although there is a price to pay, namely the loss of conformal invariance. In other words, there is now a natural scale in the problem, which determines the soliton size.

Another possibility is to add only a potential term to the action. Normally this would preclude a static configuration under the terms of Derrick's theorem. But the theorem can be side-stepped by giving the solitons angular momentum, so that the fields themselves are not static; it is now only the energy density that is independent of time. A feature of this approach is that conformal invariance is preserved, *i.e.* the model still admits solitons of arbitrary size. In many ways they are very similar to Coleman's Q-balls.[21]

For the moment the discussion will be left here, but some further remarks concerning both of these possible modifications are included in the final chapter.

Chapter VI

Low-Energy Scattering in the O(3) Model

Unfortunately, investigations of soliton scattering cannot make use of the numerical procedures developed in the previous two chapters, since they are restricted to radially symmetric energy configurations. Although the central idea of incorporating topological aspects via the Bogomolny bound could theoretically be extended to a more general scheme, the details of its implementation seem prohibitively intricate.

The aim of this chapter is to attack the question of scattering processes using a totally different approach, namely a more complete version of the "slow-motion" approximation mentioned in chapter IV (it is also sometimes called the "geodesic" approximation, for reasons that will become apparent). The technique was first suggested by Manton[15] in connection with the scattering of BPS monopoles, and it has also been applied to the interactions of critically coupled flux vortices in the abelian Higgs model. As far as its application here is concerned, it is best to think of the O(3) σ-model in the guise of the $\mathbb{C}\mathrm{P}^1$ model, *i.e.* we shall talk about a single complex field, u.

The slow-motion approximation may be applied to those theories which have both a nontrivial topology (with an associated Bogomolny bound), and static multisoliton solutions, *i.e.* no forces between isolated solitons. Often, these multisoliton solutions can be constructed explicitly. In the case of the Kähler σ-models, this was done in chapter II. To take another example, static multimonopole configurations have been found using the powerful techniques of algebraic geometry,[53] and are discussed briefly in appendix A. This is in contrast to the case of vortices, where no explicit solutions are known; but a common feature of all three examples is that there are no explicit time-dependent solutions (apart, of course, from the Lorentz boosts of static ones).

To get round this lack of time-dependent solutions, Manton suggested that, if the kinetic energy is small, the true evolution may be approximated by a sequence of static multisoliton configurations. The full field theory is truncated to a finite-dimensional dynamical system, in which the degrees of freedom are simply the parameters of the general static solution. The static solutions form a manifold, which is equipped with a natural metric coming from the kinetic part of the action, and the evolution is given by

the resulting geodesics. One can think of the static solutions as lying on a valley floor, corresponding to the minimum of the action, as given by the Bogomolny bound. Because the kinetic energy is small, the trajectories of the system can move along the valley floor but cannot climb very far up the sides.

The next section briefly reviews the approximation as applied to monopoles and vortices, while the rest of the chapter is concerned solely with the $\mathbb{C}\mathrm{P}^1$ model. Although the approach is essentially analytic, a considerable amount of numerical work will be required later on, in order to carry things through to completion.

6.1 Introduction

The geodesic approximation may be used for any particular value k of the topological charge. The evolution takes place on the manifold M_k of static charge-k solutions. In the simplest case, M_1 describes the motion at constant velocity of a single soliton, although it does not predict any Lorentz contraction, a reflection of the fact that the scheme is valid only at small velocities. In fact, here the approximation is somewhat redundant, since the same solutions are obtained *exactly* by boosting the static ones. Of more interest are the manifolds M_k where $k > 1$. Asymptotically, M_k is flat and looks like k copies of M_1. In this region the geodesics are straight lines and correspond to the motion of k well-separated solitons, each moving with constant velocity. However, the interior of M_k is, in general, curved and it is the nontrivial geometry in this region that can lead to nontrivial interactions. We are interested in tracing geodesics from the asymptotic region into the inner region and then back to the asymptotic region. Physically, this corresponds to initially well-separated solitons moving towards each other, scattering, and then moving apart once more. There is also the possibility of geodesics that form closed loops in M_k, corresponding to bound states of two or more solitons, but we shall not be concerned with these here.

In general, the evolution equations (*i.e.* the geodesic equations on M_k) take the form of n_k coupled nonlinear second-order (ordinary) differential equations, where n_k is the dimension of M_k. Things are somewhat simplified by fixing the centre of mass and also removing any overall phases, thereby effectively reducing n_k. For example, in the case of monopoles in three dimensions the full manifold of static solutions has $n_k = 4k$, but by fixing the centre of mass (three parameters) and removing an overall phase one has

effectively $n_k = 4k - 4$. Likewise, for vortices $n_k = 2k$, but fixing the centre of mass (there is no overall phase here) reduces this to $2k - 2$. Finally, for the $\mathbb{C}\mathrm{P}^1$ model, one has a full manifold of dimension

$$n_k = 4k + 2\,, \tag{6.1}$$

but fixing the centre of mass (two parameters) and the (three) overall phases gives

$$n_k = 4k - 3\,. \tag{6.2}$$

Before proceeding in any detail, we shall briefly look at the current understanding of soliton scatterings in the three models mentioned above. The simplest studies of soliton interactions correspond to $k = 2$. At first sight, vortices present the simplest calculation, with $n_2 = 2$, but progress is hampered by the absence of an explicit parametrization of the static solutions. Despite this, a qualitative picture of vortex interactions *has* been obtained.[54] The most interesting feature is that when two vortices collide head-on they emerge at 90° to their original direction of motion. This behaviour has been confirmed numerically.[51,52] Numerical studies also reveal that radiative effects are small provided that the impact velocity is less than about half the speed of light. This is a good indication that the geodesic approximation is valid for this range of velocities.

In the case of monopoles the dynamics are richer, owing to the presence of an internal relative phase. Only certain special geodesics have been identified,[55,56] but they already reveal several qualitatively different behaviours. The relative phase determines the plane of the interaction. If the initial velocities lie in this plane (type I scattering) then so does the whole evolution, with a head-on collision producing a scattering angle of 90°. On the other hand, if the initial velocities are perpendicular to the interaction plane (type II scattering) then two possible evolutions can occur, depending on the initial angular momentum, μ. When $\mu > 1$ the motion is asymptotically planar. But when $\mu \leq 1$ the monopoles always emerge along the line through the centre of mass perpendicular to the plane of the original motion. Moreover, their spatial angular momentum is converted into internal angular momentum. The physical interpretation is that the monopoles acquire equal and opposite electric charges. In other words, an interaction can convert monopoles into dyons. So far, these results have not been confirmed by numerical simulation, but at least there are analytic estimates of radiative effects in monopole scattering.[57] It is claimed that for impact velocity v (in units where the speed of light is 1) the ratio of

radiative kinetic energy to total kinetic energy is proportional to v^3. Hence it is expected that for monopoles the geodesic approximation, which ignores all radiation, is again valid for impact velocities up to about half the speed of light.

By contrast, in the $\mathbb{C}\mathrm{P}^1$ model interactions are not well understood. A small amount of work has been done,[46,47,58] and a few geodesics have been found analytically, but no systematic study has yet been made. One expects to find several qualitative differences from vortex and monopole scattering, since the theory is conformally invariant and therefore the solitons may have arbitrary size. (For both vortices and monopoles, the form of the Higgs potential provides a scale and thereby fixes their size.) To make these ideas more explicit, let us consider the time-independent solutions of the $\mathbb{C}\mathrm{P}^1$ model in a little more detail. The theory contains a single complex scalar field, u, which is a function of the space-time coordinates t, x and y. The action is given by

$$S = \iiint \frac{\partial_\mu u \, \partial^\mu \bar{u}}{(1+u\bar{u})^2} \, dx \, dy \, dt \,. \tag{6.3}$$

Here, μ labels the space-time coordinates and the space-time metric is assumed to be $\mathrm{diag}(-1,+1,+1)$. The topological charge may be written in the form

$$k = \frac{i}{2\pi} \iint \frac{\partial_x u \partial_y \bar{u} - \partial_y u \partial_x \bar{u}}{(1+u\bar{u})^2} \, dx \, dy \,. \tag{6.4}$$

It is now useful to write $z = x + iy$; then the multisoliton solutions of charge k are given by u being any rational function of z of degree k. So, for example, the most general solution with unit charge is

$$u = \lambda + \frac{\mu}{z-\nu} \,, \tag{6.5}$$

where λ, μ and ν are complex constants. By using a stereographic projection, it is seen that the target space is topologically a two-sphere, which may be oriented in internal space how ever one likes, without affecting the action. The choice of orientation corresponds to fixing the three overall phases, thereby removing three real degrees of freedom from the solution. It is conventional to choose λ to be zero, and μ to be real and positive. Physically, (6.5) then corresponds to a lump (soliton) of width μ centred at a position ν in the complex plane. One could also fix the centre of mass by setting $\nu = 0$. Then the lump would sit at the origin.

Moving on now to charge-two solutions, the most general expression for u contains ten real parameters and may be written

$$u = \alpha + \frac{2\beta z + \gamma}{z^2 + \delta z + \epsilon}\,, \tag{6.6}$$

where α, β, γ, δ and ϵ are complex constants. Again, one uses the overall phases to set α to zero, and β to be real and positive. The centre of mass is fixed by requiring that δ also vanishes. In this way, M_2 is reduced to a five-dimensional manifold, parametrized by γ and ϵ (complex) and β (real). Each point in M_2 corresponds to a field u of the form

$$u = \frac{2\beta z + \gamma}{z^2 + \epsilon}\,. \tag{6.7}$$

By analogy with the charge-one solutions, the natural physical interpretation is that there is a soliton located at each pole of u, with a width equal to the modulus of the corresponding residue. This is a good picture of things provided that the widths are small compared with the soliton separation, *i.e.* provided that we are in the asymptotic region of M_2. Explicitly, the solitons are centred at $z = \pm i\epsilon^{1/2}$ and have widths

$$\left|\frac{\gamma}{2\epsilon^{1/2}} + i\beta\right| \qquad \text{and} \qquad \left|\frac{\gamma}{2\epsilon^{1/2}} - i\beta\right| \tag{6.8}$$

respectively. However, if either of the widths (6.8) becomes large compared with $|\epsilon^{1/2}|$ then the picture of two well-separated solitons is no longer a good approximation. Although the energy density still has two maxima, the overall configuration may look quite complicated. One can take the view that the solitons distort each other when they are close together.

As the solitons interact, their sizes can change. In particular, there is the possibility that (at least) one of them could become an infinitely tall spike of zero width. Physically, this corresponds to one of the widths (6.8) becoming zero. Algebraically, it corresponds to the expression (6.7) for the field u having a common factor in the numerator and denominator, so that effectively the degree of u is no longer equal to two. As far as the geodesic approximation is concerned, the manifold M_2 has a boundary, given by

$$\gamma^2 + 4\beta^2\epsilon = 0\,, \tag{6.9}$$

and a spike occurs if a geodesic coming in from the asymptotic region hits the boundary. This situation does not seem to arise for vortices and monopoles; in these cases all the evidence accumulated so far suggests that any geodesic coming in from the asymptotic region must eventually return there.

The remainder of this chapter is arranged as follows. The next section takes the kinetic part of the action (6.3) for fields of the form (6.7) and from it derives the geodesic equations of motion. At first sight, there will be five coupled differential equations (recall from (6.2) that $n_2 = 5$), but in fact it turns out that β must remain constant throughout the evolution, in order to avoid an infinite kinetic energy. Hence there are only four coupled equations, with β now a parameter to be specified but not evolved. In essence, the geodesic evolution takes place on some member of a family of four-dimensional manifolds, labelled by β. Section 6.3 describes how to choose initial conditions for the geodesic evolution, *i.e.* how to relate β, γ, ϵ, $\dot{\gamma}$ and $\dot{\epsilon}$ to physical parameters — positions, widths, velocities and dilations. It turns out that there are hidden degrees of freedom (analogous to the relative phase in monopole scattering), which do not affect the initial motion of the solitons but can have great bearing on the eventual evolution.

Sections 6.4 and 6.5 deal with two special cases, in which the evolution is confined to a two-dimensional submanifold of M_2. In the first, β is zero, with γ and ϵ real throughout; in the second, γ is identically zero. The resulting simplifications mean that the scattering processes can be investigated analytically. Sections 6.6 and 6.7 present the numerical evolutions of some more general scatterings. The final section contains conclusions and a few general observations.

6.2 The Geodesic Equations

It seems that the full geodesic approximation is not tractable analytically; at some stage one must resort to numerical methods. The aim of this section is to formulate the problem in a way that is particularly suited to a numerical approach. Later on, we shall investigate some special geodesics using analytic techniques, but for the moment things remain completely general.

The geodesic approximation is equivalent to making an ansatz for the sigma model field of the form (6.7), with γ and ϵ being functions of time. From the general sigma model action (6.3), the kinetic energy of such configurations is seen to be

$$T = \int_0^\infty r\,dr \int_0^{2\pi} d\theta\, \frac{\left|\dot{\gamma}(z^2+\epsilon) - \dot{\epsilon}(2\beta z + \gamma)\right|^2}{\left(|z^2+\epsilon|^2 + |2\beta z + \gamma|^2\right)^2}\,, \tag{6.10}$$

where r and θ are the usual polar coordinates. For the purposes of numerical work, it is

useful to resolve γ and ϵ into their real and imaginary parts:

$$\begin{aligned} \gamma &= C + iD\,, \\ \epsilon &= E + iF\,. \end{aligned} \tag{6.11}$$

Having done this, it is not difficult to write the kinetic energy (6.10) in the form

$$T = P(\dot{C}^2 + \dot{D}^2) + Q(\dot{E}^2 + \dot{F}^2) + 2R(\dot{C}\dot{E} + \dot{D}\dot{F}) + 2S(\dot{C}\dot{F} - \dot{D}\dot{E})\,, \tag{6.12}$$

where

$$\begin{aligned} P &\equiv \int_0^\infty r\,dr \int_0^{2\pi} d\theta\, \frac{(E + r^2\cos 2\theta)^2 + (F + r^2 \sin 2\theta)^2}{\Delta^2} \\ Q &\equiv \int_0^\infty r\,dr \int_0^{2\pi} d\theta\, \frac{(C + 2\beta r\cos\theta)^2 + (D + 2\beta r \sin\theta)^2}{\Delta^2} \\ R &\equiv \int_0^\infty r\,dr \int_0^{2\pi} d\theta\, \frac{-(E + r^2\cos 2\theta)(C + 2\beta r\cos\theta) - (F + r^2 \sin 2\theta)(D + 2\beta r \sin\theta)}{\Delta^2} \\ S &\equiv \int_0^\infty r\,dr \int_0^{2\pi} d\theta\, \frac{(E + r^2\cos 2\theta)(D + 2\beta r\sin\theta) - (F + r^2 \sin 2\theta)(C + 2\beta r \cos\theta)}{\Delta^2} \end{aligned} \tag{6.13}$$

and

$$\Delta = (C + 2\beta r\cos\theta)^2 + (D + 2\beta r\sin\theta)^2 + (E + r^2\cos 2\theta)^2 + (F + r^2\sin 2\theta)^2\,. \tag{6.14}$$

In this scheme, C, D, E and F are the coordinates on the four-dimensional manifold M_2. The quantities P, Q, R and S are functions of C, D, E and F, but do not depend on time. The natural metric on M_2 coming from (6.12) is

$$ds^2 = P(dC^2 + dD^2) + Q(dE^2 + dF^2) + 2R(dC\,dE + dD\,dF) + 2S(dC\,dF - dD\,dE)\,. \tag{6.15}$$

The equations (6.12) and (6.15) simply provide two alternative ways of viewing evolution in the low-energy approximation. Since every configuration of the form (6.7) has the same

potential energy, the kinetic energy (6.12) may be taken as the Lagrangian for a restricted $\mathbb{C}\mathrm{P}^1$ model, in which C, D, E and F are the degrees of freedom. The Euler-Lagrange equations derived from (6.12) are then precisely the geodesic equations associated with the metric (6.15).

Even at this stage, there are symmetries apparent in (6.12) which reveal (at least) three totally geodesic submanifolds of dimension two. First, T is invariant under the transformation $D \to -D$ and $F \to -F$. To see this, make a change of variable $\theta \to -\theta$ in (6.13) and (6.14). Hence, taking $D = F = 0$ (*i.e.* γ and ϵ both real) specifies a totally geodesic submanifold of M_2. In other words, if the initial conditions are such that both γ and ϵ (and their time derivatives) are real at $t = 0$, then they remain real throughout the evolution. Second, T is also invariant under $C \to -C$ and $F \to -F$, as seen by taking $\theta \to \pi - \theta$ in (6.13) and (6.14). The associated totally geodesic submanifold has γ pure imaginary and ϵ pure real. Finally, T is again invariant under $C \to -C$ and $D \to -D$ (take $\theta \to \pi + \theta$), corresponding to a submanifold with γ identically zero. There are almost certainly other examples, but they do not appear to have such simple parametrizations.

It is interesting to note that if ϵ is real throughout (as in two of the geodesic submanifolds mentioned above) then the solitons are constrained to lie on either the real (if $\epsilon < 0$) or the imaginary (if $\epsilon > 0$) axis in the complex plane. In particular, if ϵ were to evolve from a negative value to a positive one, or vice-versa, then one would see an interaction with a scattering angle of 90°. This is the first hint that the $\mathbb{C}\mathrm{P}^1$ solitons have scattering properties in common with other nonintegrable objects such as vortices and monopoles.

Returning to the general case, we conclude this section with a few remarks concerning the evolution equations derived from (6.12) or (6.15). In the most convenient form for numerical implementation they are

$$\begin{aligned}
\ddot{C} &= \frac{Q\alpha_C - R\alpha_E - S\alpha_F}{R^2 + S^2 - PQ}\,, \\
\ddot{D} &= \frac{Q\alpha_D + S\alpha_E - R\alpha_F}{R^2 + S^2 - PQ}\,, \\
\ddot{E} &= \frac{S\alpha_D + P\alpha_E - R\alpha_C}{R^2 + S^2 - PQ}\,, \\
\ddot{F} &= \frac{P\alpha_F - S\alpha_C - R\alpha_D}{R^2 + S^2 - PQ}\,.
\end{aligned} \tag{6.16}$$

In (6.16), α_C, α_D, α_E and α_F contain only first derivatives. More precisely, they are quadratic in the time derivatives of C, D, E and F, and linear in the derivatives of P, Q, R and S with respect to C, D, E and F. It is not particularly instructive to write them all out explicitly, but for example

$$\begin{aligned} 2\alpha_C = {} & \frac{\partial P}{\partial C}\left(\dot{C}^2 - \dot{D}^2\right) + \left(2\frac{\partial R}{\partial E} - \frac{\partial Q}{\partial C}\right)\dot{E}^2 + \left(2\frac{\partial S}{\partial F} - \frac{\partial Q}{\partial C}\right)\dot{F}^2 \\ & + 2\left(\frac{\partial R}{\partial D} + \frac{\partial S}{\partial C}\right)\dot{D}\dot{E} + 2\left(\frac{\partial S}{\partial D} - \frac{\partial R}{\partial C}\right)\dot{D}\dot{F} + 2\left(\frac{\partial R}{\partial F} + \frac{\partial S}{\partial E}\right)\dot{E}\dot{F} \\ & + 2\frac{\partial P}{\partial D}\dot{C}\dot{D} + 2\frac{\partial P}{\partial E}\dot{C}\dot{E} + 2\frac{\partial P}{\partial F}\dot{C}\dot{F}\,. \end{aligned} \tag{6.17}$$

The scheme outlined above has been implemented in FORTRAN on the Amdahl mainframe in Durham, making extensive use of the NAG library routines (see the program listing in appendix C). At first sight, in addition to evolving the system (6.16), one must calculate P, Q, R, S and all their derivatives using the expressions (6.13). This represents a total of 20 two-dimensional numerical integrations merely to calculate the right hand sides of (6.16). However, this number is reduced to 16 by noting that there are four linear relations between the various derivatives:

$$\begin{aligned} \frac{\partial Q}{\partial C} - \frac{\partial R}{\partial E} - \frac{\partial S}{\partial F} &= 0\,, \\ \frac{\partial Q}{\partial D} - \frac{\partial R}{\partial F} + \frac{\partial S}{\partial E} &= 0\,, \\ \frac{\partial P}{\partial E} - \frac{\partial R}{\partial C} + \frac{\partial S}{\partial D} &= 0\,, \\ \frac{\partial P}{\partial F} - \frac{\partial R}{\partial D} - \frac{\partial S}{\partial C} &= 0\,. \end{aligned} \tag{6.18}$$

Finally, there is a word of warning. From (6.16), it is clear that a numerical implementation could get into difficulty if $R^2 + S^2 - PQ = 0$. However, it seems far from easy to see what is the corresponding condition on C, D, E and F. Suffice to remark here that extensive numerical work has not encountered this problem so far.

6.3 Initial Data and Physical Parameters

In this section we move on to the question of choosing initial data for the geodesic evolution. Since the geodesic equations are second order, one must specify γ, ϵ, $\dot{\gamma}$ and $\dot{\epsilon}$ at time $t = 0$, in addition to the label β (recall that β cannot vary with time). To do this, we shall relate them to the physical parameters shown in figure 6.1. Instead of resolving γ and ϵ into real and imaginary parts, as in the previous section, here it is more useful to write them in the form

$$\gamma = Ae^{i\sigma}\,, \quad \epsilon^{1/2} = Be^{i\tau}\,, \tag{6.19}$$

where A and B are positive (or possibly zero).

The initial positions of the solitons are given by a dimensionless impact parameter J and an overall scale λ, which is chosen to be much greater than unity, so that the initial configuration lies in the asymptotic region of the manifold M_2. Their widths are taken to be w_+ and w_-, with $w_+ \geq w_-$. It is useful to define w and δ by

$$w_\pm = w \pm \delta\,, \quad (0 \leq \delta < w)\,. \tag{6.20}$$

The link with β, γ and ϵ comes from the asymptotic picture described in the introduction, namely that in the asymptotic region of M_2 the soliton positions are $\pm i\epsilon^{1/2}$ and their widths are

$$\left|\frac{\gamma}{2\epsilon^{1/2}} + i\beta\right| \qquad \text{and} \qquad \left|\frac{\gamma}{2\epsilon^{1/2}} - i\beta\right|\,. \tag{6.21}$$

Without loss of generality, one can take the solitons to lie in the top-right and bottom-left quadrants of the xy-plane. Furthermore, it is clear from (6.21) that changing σ by π simply has the effect of interchanging the two solitons. Therefore, in what follows we can assume that

$$-\pi/2 \leq \tau < 0 \quad \text{and} \quad 0 \leq (\sigma - \tau) < \pi\,. \tag{6.22}$$

Matching the positions and widths to γ and ϵ yields

$$B = \lambda w\,, \tag{6.23}$$

$$\tau = \cos^{-1}(J/\lambda)\,, \tag{6.24}$$

$$A = 2\lambda w\sqrt{w^2 + \delta^2 - \beta^2}\,, \tag{6.25}$$

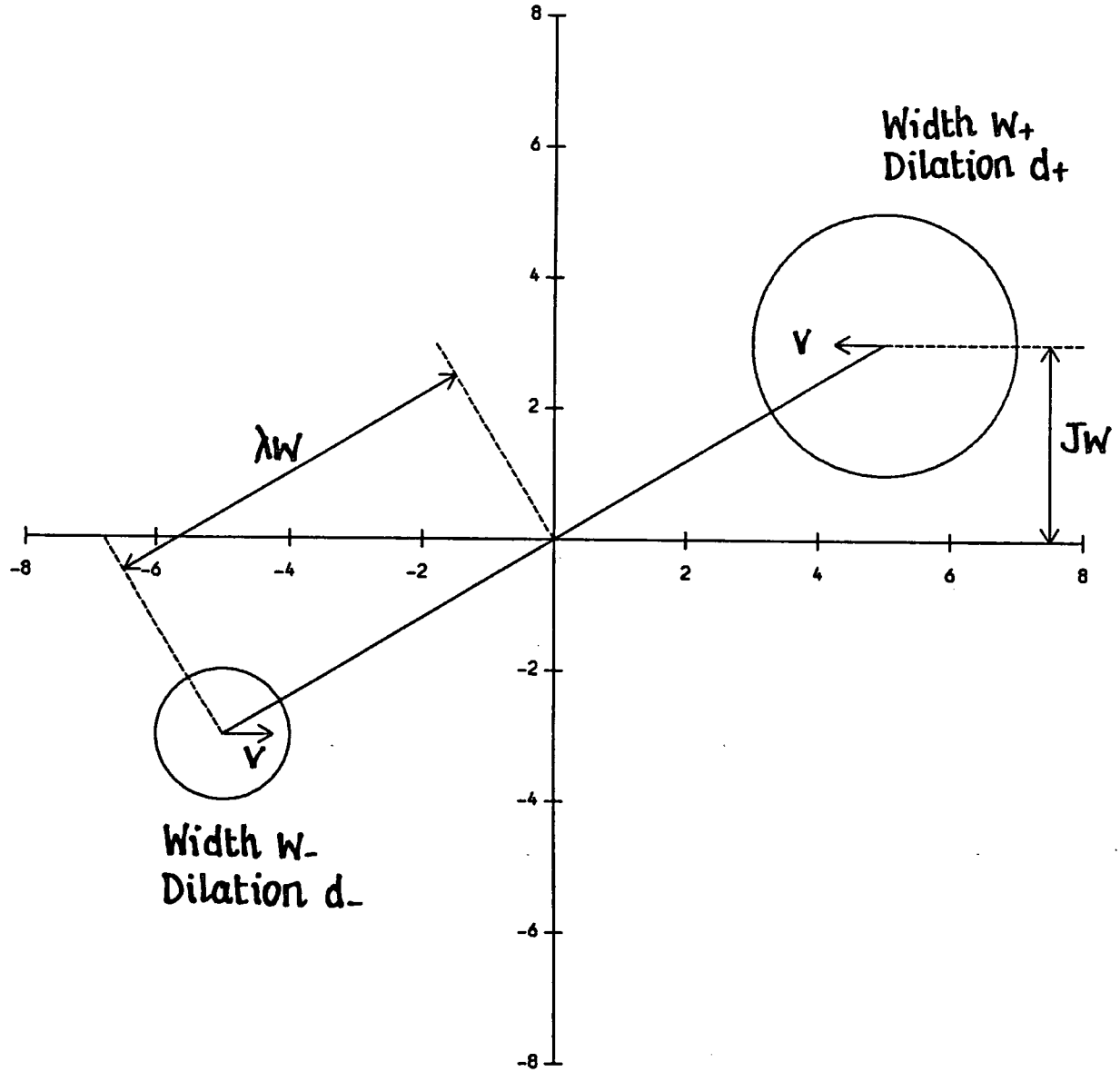


FIG 6.1. The physical parameters used to specify initial conditions for the geodesic evolution. The quantity w is the average of the individual widths $w_{\pm}$.

$$w\delta = \beta\sqrt{w^2+\delta^2-\beta^2}\,\sin(\sigma-\tau)\,. \tag{6.26}$$

At this point, one should clarify the role of β. In addition to the four physical parameters J, λ, w and δ, one must also specify β in order to determine γ and ϵ. But what values can β take? Clearly, β^2 cannot be greater than $w^2+\delta^2$, to ensure that the square roots in (6.25) and (6.26) are real. However, since $0 \leq \sin(\sigma-\tau) \leq 1$, (6.26) constrains β a little further. In fact, one finds that

$$\delta \leq \beta \leq w\,. \tag{6.27}$$

The natural interpretation is that β is a "free" parameter, subject only to the constraint (6.27). It does not affect the initial energy density, but will, in general, affect the form of the interaction.

Turning now to the specification of $\dot{\gamma}$ and $\dot{\epsilon}$, the physical parameters to be matched here are the initial velocity, v, and the initial dilations $d_\pm$, defined by

$$d_\pm = \left.\frac{dw_\pm}{dt}\right|_{t=0}\,. \tag{6.28}$$

Using the Lorentz invariance of the model, we are free to take the initial velocity to be parallel to the x-axis, as shown in figure 6.1. Performing the algebra yields

$$\dot{B} = v\sin\tau\,, \tag{6.29}$$

$$\dot{\tau} = Jv/\lambda^2 w\,, \tag{6.30}$$

$$\sqrt{w^2+\delta^2-\beta^2}\left(\dot{A}-2v\sin\tau\sqrt{w^2+\delta^2-\beta^2}\right) = \lambda w D_+\,, \tag{6.31}$$

$$\begin{aligned}\beta\sin(\sigma-\tau)&\left(\dot{A}-2v\sin\tau\sqrt{w^2+\delta^2-\beta^2}\right)\\ &+2\beta\lambda w\sqrt{w^2+\delta^2-\beta^2}\,(\dot{\sigma}-\dot{\tau})\cos(\sigma-\tau) = \lambda w D_-\,,\end{aligned} \tag{6.32}$$

where

$$\begin{aligned}D_+ &\equiv w_+ d_+ + w_- d_-\,,\\ D_- &\equiv w_+ d_+ - w_- d_-\,.\end{aligned} \tag{6.33}$$

Note that (6.26) has in general two solutions for σ, which lead to two different values of $\dot{\sigma}$ from (6.32). However, using the symmetry of the metric under $\sigma \to -\sigma$ and $\tau \to -\tau$, it is

not difficult to see that the two possibilities correspond to essentially the same interaction. Without loss of generality, one can take $\cos(\sigma - \tau) \geq 0$, *i.e.* (6.22) can be modified so that

$$0 \leq (\sigma - \tau) \leq \pi/2 \,. \tag{6.34}$$

To sum up, there are seven physical parameters, namely J, λ, w, δ, v and $d_\pm$. The free parameter β must be chosen to lie between δ and w; then the values of γ and ϵ at $t = 0$ are determined by (6.23)–(6.26) in terms of J, λ, w and δ. Likewise, v and $d_\pm$ determine $\dot{\gamma}$ and $\dot{\epsilon}$ via (6.29)–(6.32). On the face of it, $\dot{\gamma}$ and $\dot{\epsilon}$ should correspond to four physical parameters, instead of just three, but one is absorbed by taking the initial velocity parallel to the x-axis. Note that if v, d_+ and d_- are all scaled by the same factor then the geodesic that is traced out is unchanged, but it will be traced out at a different rate. As far as the geodesics are concerned, time is simply an affine parameter.

We now have a general picture of things, but it is worth mentioning some special cases, in which extra free parameters emerge in return for a physical constraint on the relative dilations. They occur when β takes a value at the end of its range, *i.e.* when it is equal to either δ or w. In addition, the situation is slightly different depending on whether δ is equal to zero. So in all there are four special cases, which we shall consider in turn.

(i) $0 = \delta = \beta < w$

This is the only case in which $\beta = 0$, and is the situation previously studied by Ward.[46] From (6.32) one has the physical constraint that $d_+ - d_- = 0$. In return, there appear to be two free parameters, namely σ and $\dot{\sigma}$, but σ may be absorbed into an overall phase (since $\beta = 0$) and is therefore irrelevant. On the other hand, we shall see later that the choice of $\dot{\sigma}$ has a great bearing on the eventual interaction. Equations (6.25) and (6.31) simplify to give

$$\begin{aligned} A &= 2\lambda w^2 \,, \\ \dot{A} &= 2w(\lambda d_+ + v \sin\tau) \,. \end{aligned} \tag{6.35}$$

(ii) $0 = \delta < \beta = w$

This is the only case in which $A = 0$. The physical constraint is $d_+ + d_- = 0$. It appears that $\dot{\sigma}$ is a free parameter, but in fact this is spurious: since $A = 0$, $\dot{\sigma}$ affects neither γ nor $\dot{\gamma}$. The free parameter comes from the fact that $\dot{A}$ and σ are determined by a single equation, namely

$$\dot{A}\sin(\sigma - \tau) = 2\lambda w d_+ \,. \tag{6.36}$$

(iii) $0 < \delta = \beta < w$

Here the physical constraint is $d_+ - d_- = 0$ and the free parameter is $\dot{\sigma}$. The equations for σ, A and $\dot{A}$ simplify to give

$$\begin{aligned} \sigma &= \tau + \pi/2\,, \\ A &= 2\lambda w^2\,, \\ \dot{A} &= 2w(\lambda d_+ + v\sin\tau)\,. \end{aligned} \tag{6.37}$$

(iv) $0 < \delta < \beta = w$

Here the physical constraint is $d_+ + d_- = 0$ and the free parameter is again $\dot{\sigma}$. The equations for σ, A and $\dot{A}$ become

$$\begin{aligned} \sigma &= \tau + \pi/2\,, \\ A &= 2\lambda w\delta\,, \\ \dot{A} &= 2(\lambda w d_+ + v\delta\sin\tau)\,. \end{aligned} \tag{6.38}$$

It is also instructive to look at the physical consequences of restricting the evolution to each of the three totally geodesic submanifolds identified in the previous section. It turns out that there are strong connections with the four special cases discussed above.

(a) γ real, ϵ real

In this case, $\sigma = 0$ and $\tau = -\pi/2$, with $\dot{\sigma} = \dot{\tau} = 0$. Physically, this means that the collision is head-on. In addition, from equation (6.26) β must be equal to either δ or w, *i.e.* we have one of the four special cases already discussed, but without the extra free parameters. Explicitly, the initial conditions are as follows.

$$\text{If } \beta = \delta \text{ then } \quad \begin{cases} \gamma = 2\lambda w^2 \,, \\ \dot{\gamma} = 2w(\lambda d_+ - v) \,, \\ \epsilon = -\lambda^2 w^2 \,, \\ \dot{\epsilon} = 2\lambda w v \,, \end{cases} \tag{6.39}$$

together with the constraint $d_+ - d_- = 0$.

$$\text{If } \beta = w \text{ then } \quad \begin{cases} \gamma = 2\lambda w \delta \,, \\ \dot{\gamma} = 2(\lambda w d_+ - \delta v) \,, \\ \epsilon = -\lambda^2 w^2 \,, \\ \dot{\epsilon} = 2\lambda w v \,, \end{cases} \tag{6.40}$$

together with the constraint $d_+ + d_- = 0$.

(b) γ pure imaginary, ϵ real

In this case, $\sigma = \tau = -\pi/2$ and $\dot{\sigma} = \dot{\tau} = 0$. Again the collision is head-on; and there are also two further physical consequences. From (6.26), $\delta = 0$ and from (6.32), $d_+ = d_-$. Provided that $\beta \neq w$, the initial conditions are given by

$$\begin{aligned} \gamma &= -2i\lambda w \sqrt{w^2 - \beta^2} \,, \\ \dot{\gamma} &= -2i \left(\frac{\lambda w^2 d_+}{\sqrt{w^2 - \beta^2}} - v\sqrt{w^2 - \beta^2} \right) , \\ \epsilon &= -\lambda^2 w^2 \,, \\ \dot{\epsilon} &= 2\lambda w v \,, \end{aligned} \tag{6.41}$$

while if $\beta = w$ then one has case (ii) as discussed above (with $d_+ = d_- = 0$ and with $\dot{\gamma}$ arbitrary, subject to the requirement that it be pure imaginary). Note also that $\beta = 0$ gives essentially the same evolution as (6.39), since they differ only by an overall phase.

(c) γ identically zero

This is another subcase of case (ii). Although the collision need not be head-on, the other physical constraints are quite severe. In particular, since $\dot{A} = 0$, (6.31) and (6.32) imply that $d_+ = d_- = 0$. As far as γ and ϵ are concerned, one must simply have initial conditions with $\gamma = \dot{\gamma} = 0$.

6.4 Analytic Results : γ and ϵ both real, $\beta = 0$

Having obtained a complete classification of the initial configurations, the next two sections look in more detail at two families of evolutions that may be understood analytically. They both correspond to motion on one of the totally geodesic submanifolds identified earlier. In this section, γ and ϵ are taken to be real throughout and β is set to zero, giving a particular example of (6.39) in which the initial widths of the solitons are both equal to w. As noted before, the initial dilations also have the same value, and this will be denoted by d.

The most useful approach seems to be that adopted by Ward,[46] which parametrizes γ and ϵ using two variables, R and ψ, where $R \geq 0$ and $0 \leq \psi \leq \pi$. Explicitly,

$$\begin{aligned} \gamma &= R \sin\psi \,, \\ \epsilon &= R \cos\psi \,. \end{aligned} \tag{6.42}$$

In terms of physical parameters, the initial data for R and ψ (denoted by a subscript 0) are given by

$$\begin{aligned} R_0 &= \lambda w^2 \sqrt{4+\lambda^2} \,, \\ \psi_0 &= \tan^{-1}(-2/\lambda) \,, \\ \dot{R}_0 &= \frac{2w}{\sqrt{4+\lambda^2}} \left(2\lambda d - v(2+\lambda^2)\right) \,, \\ \dot{\psi}_0 &= \frac{-2}{w(4+\lambda^2)} \left(\lambda d + v\right) . \end{aligned} \tag{6.43}$$

Ward showed that the kinetic energy of the system may be concisely written in terms of

elliptic integrals:

$$T = \pi\big(\xi(\psi)R^{-1}\dot{R}^2 + \mu(\psi)\dot{R}\dot{\psi} + \nu(\psi)R\dot{\psi}^2\big)\,, \tag{6.44}$$

where

$$\begin{aligned}\xi(\psi) &= \tfrac{1}{2}E(\cos\psi)\,,\\ \mu(\psi) &= \tan\psi\big(K(\cos\psi) - E(\cos\psi)\big)\,,\\ \nu(\psi) &= K(\cos\psi) - \tfrac{1}{2}E(\cos\psi)\,,\end{aligned} \tag{6.45}$$

with K and E the complete elliptic integrals of the first and second kind, respectively. The evolution equation for R may be integrated twice and that for ψ integrated once to give

$$\begin{aligned}R &= T\big((t-\alpha)^2 + Q^2\big)/(4\pi\xi)\,,\\ \dot{\psi}^2 &= Q^2\Big/\Big(I(\psi)\big((t-\alpha)^2 + Q^2\big)\Big)^2\,,\end{aligned} \tag{6.46}$$

where Q (≥ 0) and α are real constants having the dimensions of time, and where

$$I(\psi) \equiv \frac{\sqrt{4\nu\xi - \mu^2}}{4\xi}\,. \tag{6.47}$$

The function $I(\psi)$ is drawn in figure 6.2. It is always positive, reaching a minimum value of $\frac{1}{2}$ when $\psi = \frac{\pi}{2}$. It tends to infinity as ψ approaches either 0 or π, but does so logarithmically, with the result that the area under the curve is finite. In fact

$$\int_0^\pi I(\psi) \approx 2.3162\,. \tag{6.48}$$

The evolution contains three real constants, namely T, Q and α. (Note that T is conserved because all points of the manifold M_2 correspond to configurations with the same potential energy, and so conservation of total energy becomes conservation of kinetic energy.) All three constants are easily determined in terms of the initial data using (6.44) and (6.46):

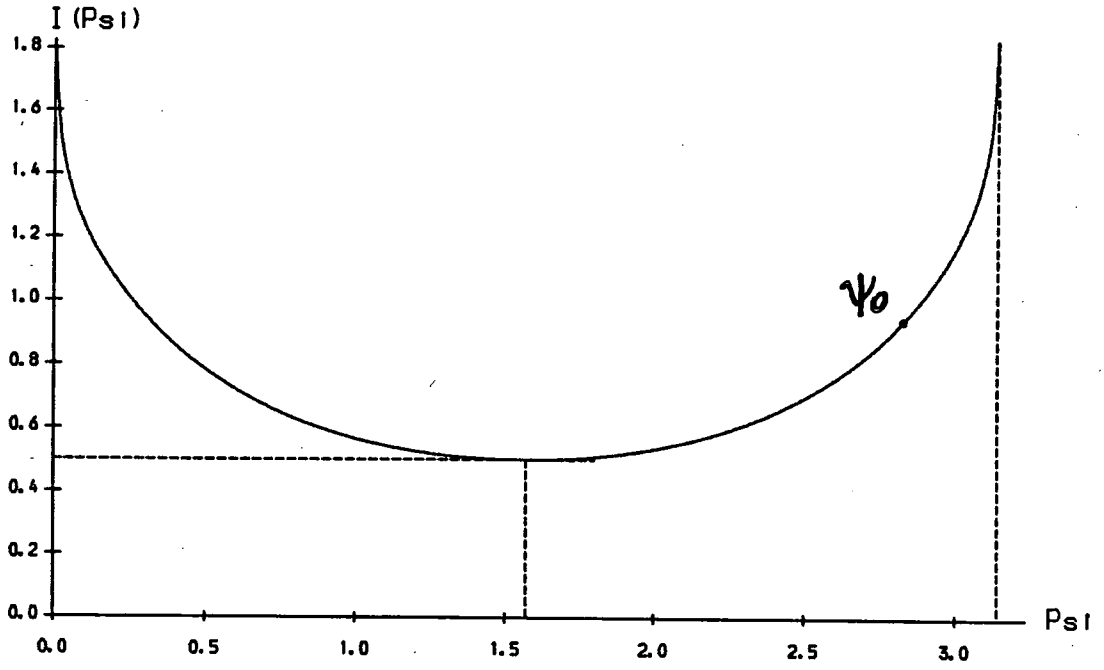


FIG 6.2. The function $I(\psi)$.

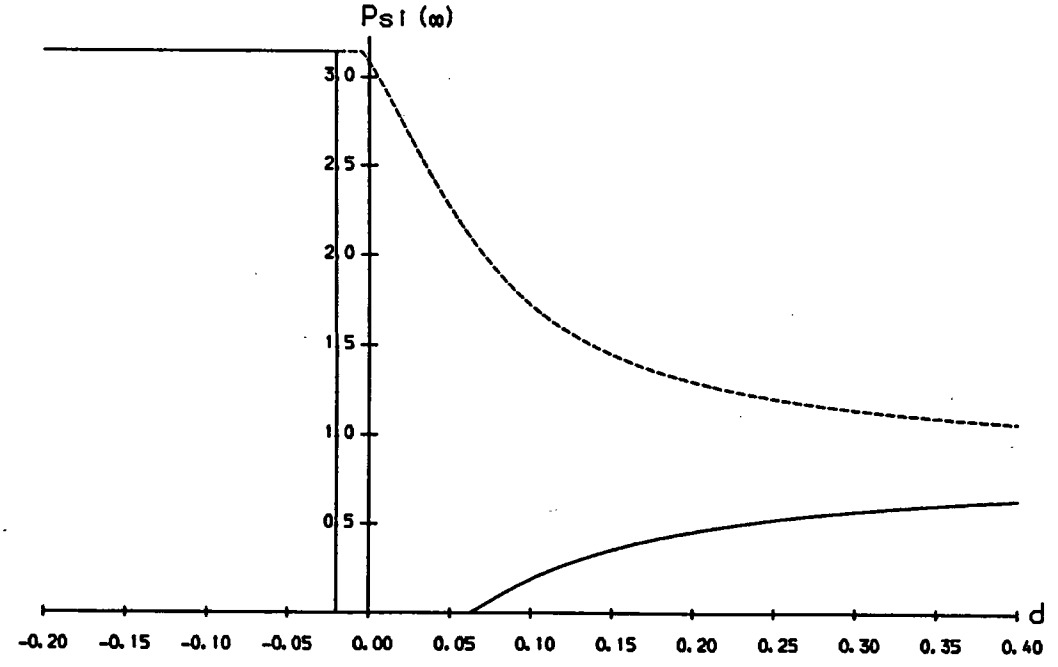


FIG 6.3. The variation of ψ_∞ with d for $w = 1$, $\lambda = 5$ and $v = \pm 0.1$ (solid and broken lines respectively).

$$
\begin{aligned}
T &= \pi\left(\xi_0 R_0^{-1}\dot{R}_0^2 + \mu_0\dot{R}_0\dot{\psi}_0 + \nu_0 R_0\dot{\psi}_0^2\right), \\
Q &= 4\pi\xi_0 R_0 I(\psi_0)|\dot{\psi}_0| \,/\, T, \\
\alpha &= -\pi(\mu_0\dot{\psi}_0 R_0 + 2\xi_0\dot{R}_0) \,/\, T.
\end{aligned}
\tag{6.49}
$$

The key to understanding the evolution is the variation of ψ. It starts out at a value close to π and as time evolves it varies monotonically. It may either increase or decrease, depending on the sign of $\dot{\psi}_0$, but given $\dot{\psi}_0$ there are always two possibilities.

$$
\text{If } \dot{\psi}_0 < 0 \text{ then } \psi \text{ may } \begin{cases} \text{become zero at some time } t = t_p. \\ \text{tend to some limit } \psi_\infty \text{ between 0 and } \psi_0. \end{cases}
$$

$$
\text{If } \dot{\psi}_0 > 0 \text{ then } \psi \text{ may } \begin{cases} \text{become equal to } \pi \text{ at some time } t = t_p. \\ \text{tend to some limit } \psi_\infty \text{ between } \psi_0 \text{ and } \pi. \end{cases}
$$

It is easy to translate the variation of ψ into information about physical processes. First, if ψ becomes either 0 or π then the solitons become spikes. Second, if ψ passes through $\pi/2$ then the solitons scatter at 90°. Therefore the solitons may scatter at right angles and then become spiky (corresponding to ψ decreasing to zero) or they may become spiky *before* scattering (corresponding to ψ increasing to π). But ψ may also tend to a limit away from 0 and π, and this leads to further, qualitatively different possibilities. Suppose that as $t \to \infty$, $\psi \to \psi_\infty$, not equal to 0 or π. Then

$$
\begin{aligned}
\gamma &\to kt^2 \sin\psi_\infty, \\
\epsilon &\to kt^2 \cos\psi_\infty,
\end{aligned}
\tag{6.50}
$$

where k is a positive constant. Physically, (6.50) corresponds to two solitons moving apart at constant velocity and expanding at a constant rate. So it is possible for the solitons to scatter and then enter an expanding phase, or to expand without scattering. In these cases they never become spikes, *i.e.* the corresponding geodesic never hits the boundary of the manifold M_2.

It only remains to match the various possibilities to different choices of initial conditions, or equivalently to different values of T, Q and α. First, suppose that $\dot{\psi}_0$ is negative,

i.e. ψ decreases monotonically. From (6.46), ψ reaches zero provided

$$\int_0^{\psi_0} I(\psi)\,d\psi \;<\; \frac{\pi}{2} + \tan^{-1}\Big(\frac{\alpha}{Q}\Big)\,. \tag{6.51}$$

In this case the solitons become spikes after a time

$$t_p \;=\; \alpha \;+\; Q\,\tan\left(\int_0^{\psi_0} I(\psi)\,d\psi - \tan^{-1}\Big(\frac{\alpha}{Q}\Big)\right) \tag{6.52}$$

at a distance

$$y_p \;=\; Q\sqrt{\frac{T}{2\pi}}\,\sec\left(\int_0^{\psi_0} I(\psi)\,d\psi - \tan^{-1}\Big(\frac{\alpha}{Q}\Big)\right) \tag{6.53}$$

from the origin along the y-axis. On the other hand, if (6.51) does not hold then ψ tends to a limiting value ψ_∞ given by

$$\int_{\psi_\infty}^{\psi_0} I(\psi)\,d\psi \;=\; \frac{\pi}{2} + \tan^{-1}\Big(\frac{\alpha}{Q}\Big)\,. \tag{6.54}$$

The corresponding equations in the case $\dot{\psi}_0 > 0$ are very similar. One simply has to change the ranges of integration to $[\psi_0, \pi]$ in (6.51)–(6.53) and to $[\psi_0, \psi_\infty]$ in (6.54). Now (6.51) becomes the condition for ψ to reach π.

As an illustration of the way in which the various possibilities fit together, consider changing the initial dilation while keeping the other parameters fixed. The solid line in figure 6.3 shows the variation of ψ_∞ with d when

$$w = 1\,, \quad \lambda = 5\,, \quad v = 0.1\,. \tag{6.55}$$

The broken line shows the effects of changing v to -0.1, *i.e.* having the solitons initially moving apart. This does not strictly have a place in an investigation of scattering processes, but is included for the sake of completeness. Concentrating on the solid line, there

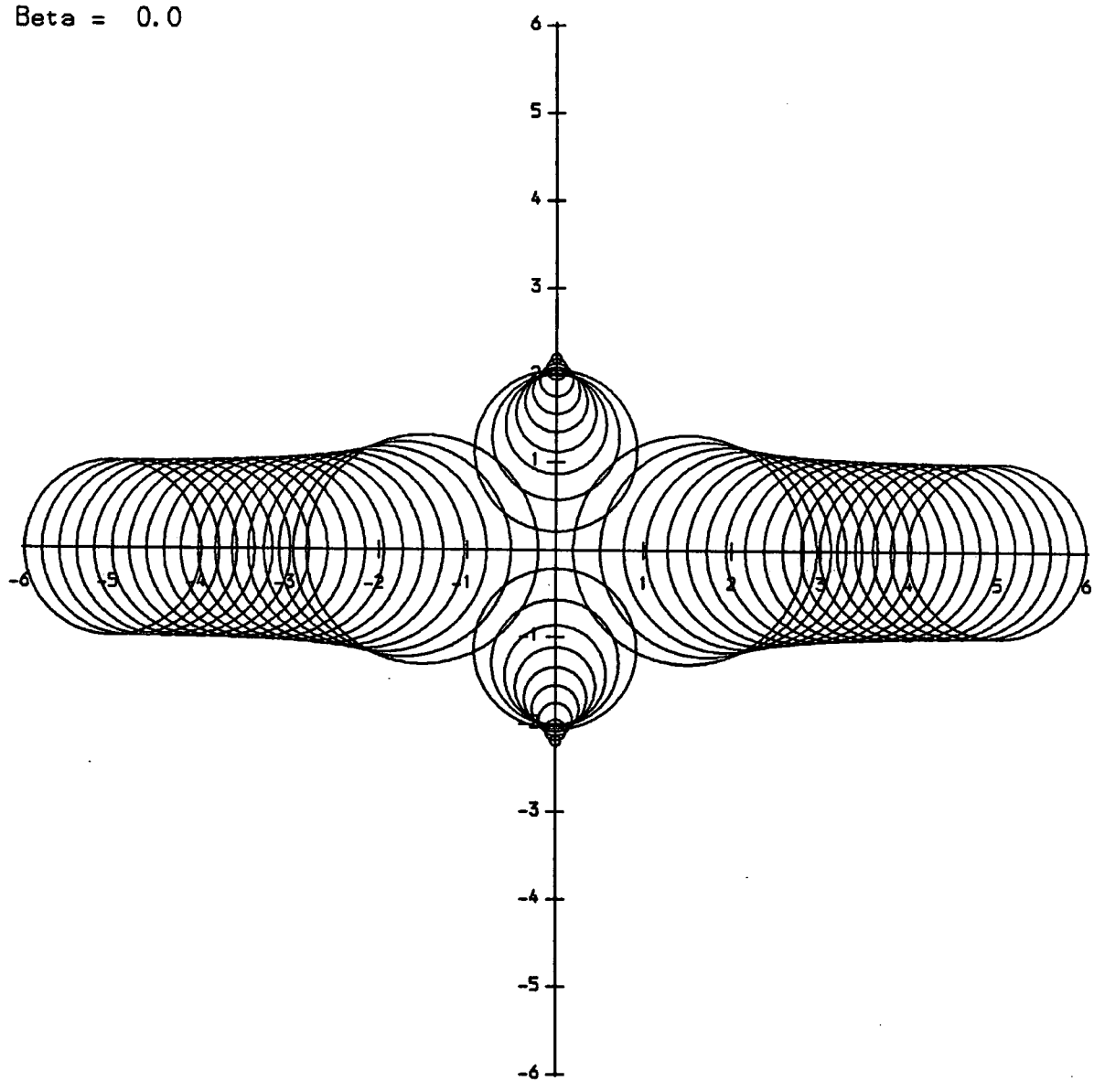


FIG 6.4. Scattering process for $w = 1$, $\lambda = 5$, $v = 0.1$ and $d = 0$.

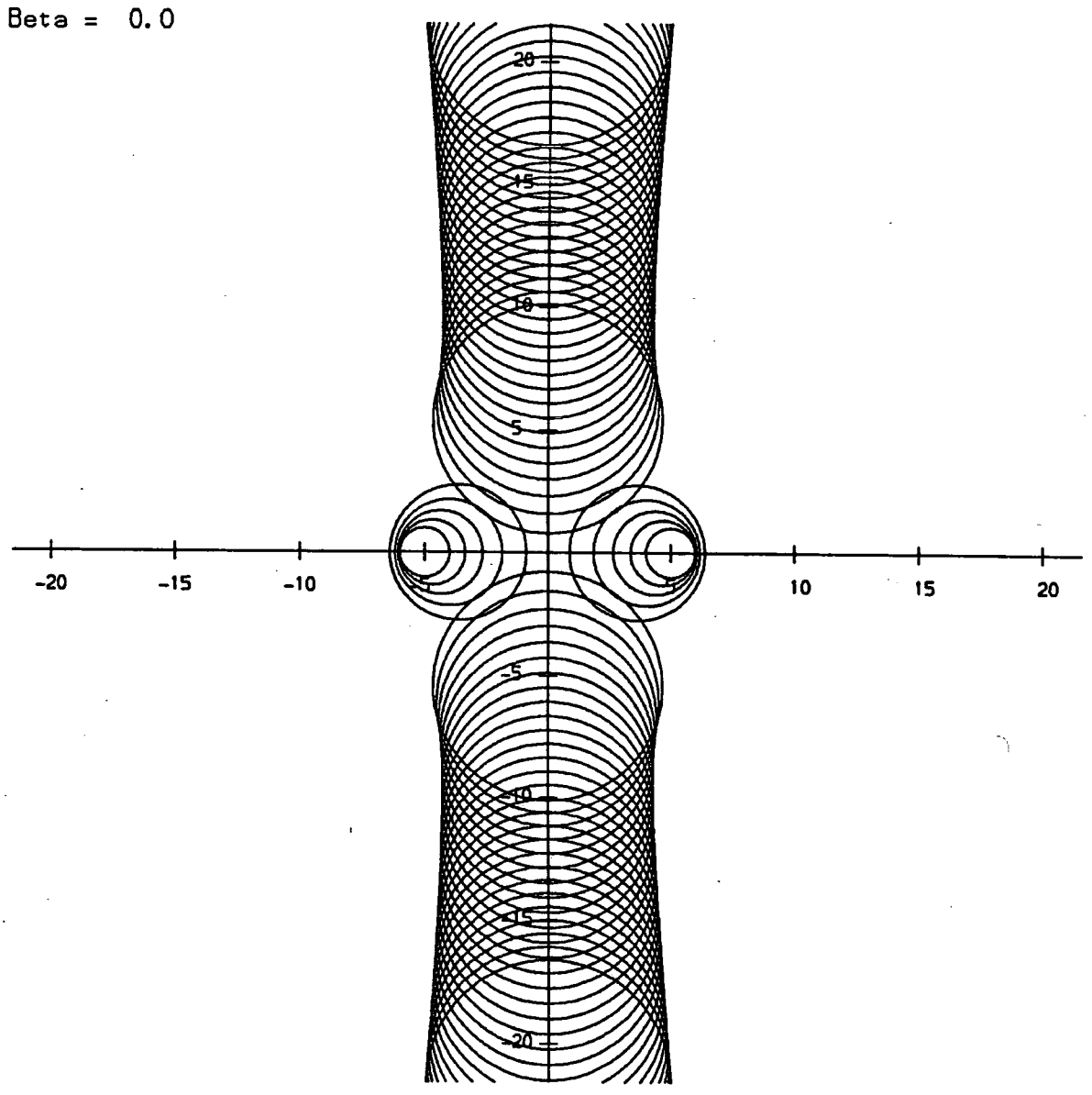


FIG 6.5. Scattering process for $w = 1$, $\lambda = 5$, $v = 0.1$ and $d = 0.1$.

are three qualitatively different behaviours.

$$\begin{aligned} &\text{For } d < -0.02\,, && \psi_\infty = \pi\,. \\ &\text{For } d \in (-0.02, 0.06158)\,, && \psi_\infty = 0\,. \\ &\text{For } d > 0.06158\,, && \psi_\infty \in (0, \pi)\,. \end{aligned}$$

These three regions correspond respectively to the solitons becoming spiky before scattering, becoming spiky after scattering and reaching an expanding phase after scattering. When $v > 0$ they cannot reach an expanding phase without scattering. When $d = -0.02$ they become spikes at precisely the moment they both reach the origin, and when $d = 0.06158$ they scatter but neither shrink nor expand. In the limit $d \to \infty$ it can be shown that ψ_∞ approaches 0.8403. However, it is questionable whether one should consider very large dilations, because in these cases the solitons eventually expand to cover the whole plane, *i.e.* the corresponding geodesic never returns to the asymptotic region of the manifold M_2. If the values of w or λ are altered then the boundaries between the three regions will move; but the general pattern is always the same. Moreover, $d = 0$ always lies in the middle region.

The evolutions for $d = 0$ and $d = 0.1$ are shown schematically in figures 6.4 and 6.5. The circles are drawn in pairs at equally spaced values of time. The centres of the circles represent the soliton positions, and the radii their widths, as given by the asymptotic picture (6.21). So, for example, when $t = 0$ two circles of unit radius are drawn at the points $(\pm 5, 0)$. As the system evolves, one can follow the soliton trajectories together with the associated changes in size. If at any particular time the pair of circles intersect (for example, during the interaction itself) then they are not drawn, first because the asymptotic picture is not reliable in such situations and second because it would make the figures unnecessarily complicated. It is worth remarking that the figures were generated using the numerical scheme set up in section 6.2; they are in excellent agreement with the analytic treatment developed above, and so provide evidence that the computer implementation is working well. For example, with reference to figure 6.4, (6.53) predicts that the solitons shrink to spikes at a distance from the origin of 2.325.

Finally, it is interesting to consider the limit $\lambda \to \infty$ in the case when $d = 0$. As remarked earlier, the solitons will scatter, then become spiky, and intuitively one does not expect the value of λ to affect things very much. As $\lambda \to \infty$, $\psi_0 \to \pi$ and using the

asymptotic properties of the elliptic functions one finds

$$Q \to \frac{w}{v}\sqrt{2\ln(2\lambda)}\,, \qquad \alpha \to \frac{\lambda w}{v}\,. \tag{6.56}$$

Therefore, recalling (6.48),

$$\int_0^{\psi_0} I(\psi)\,d\psi \;-\; \tan^{-1}\!\left(\frac{\alpha}{Q}\right) \;\to\; 2.3162 - \frac{\pi}{2} \;\approx\; 0.74544\,, \tag{6.57}$$

and then (6.52) becomes

$$t_p \approx \alpha + (0.92312)Q\,. \tag{6.58}$$

The fact that the coefficient of Q in (6.58) is close to unity, together with the much weaker dependence of Q than α on λ suggests the following physical interpretation. Roughly speaking, α is the time taken for the solitons to reach the interaction region, and Q is the time taken from the interaction to the eventual collapse into spikes. As λ increases, α increases linearly but the increase in Q is much slower.

6.5 Analytic Results : γ identically zero

A second family of analytically tractable evolutions consists of those which take place on the two-dimensional submanifold of M_2 that is obtained by setting γ to zero. This submanifold will be denoted by Σ. Essentially, one is taking the field u to be of the form

$$u = \frac{2\beta z}{z^2 + \epsilon}\,, \tag{6.59}$$

with the time dependence of u now wholly contained in ϵ. From (6.21), it is seen that each soliton has initial width β. Not only that, but in the asymptotic region the soliton widths are *always* equal to β, *i.e.* scattering processes lead to no long-term change in width. However, if one moves away from the asymptotic picture and looks at the potential energy density arising from (6.59) in the *inner* region of M_2 (where $|\epsilon^{1/2}|$ is no longer large compared to β), then one finds that the soliton widths are now less than β. In other words, the width of each soliton decreases as they approach each other, then eventually returns to its original value as they move apart. The only exception to this general behaviour occurs if ϵ ever becomes zero, at which point the solitons become spikes. The point $\epsilon = 0$ is the intersection of Σ with the boundary of M_2.

Getting down to details, it is useful to parametrize ϵ using two variables, R and ϕ, where $R \geq 0$ and $0 \leq \phi < 2\pi$:

$$\epsilon = Re^{i\phi} . \tag{6.60}$$

Hopefully this notation will not lead to confusion with previous parametrizations. From (6.10), the kinetic energy of the system may be written in the form

$$T = 4\beta^2 f(R)\,(\dot{R}^2 + R^2\dot{\phi}^2)\,, \tag{6.61}$$

where

$$f(R) \equiv \int_0^\infty dr \int_0^{2\pi} d\theta \; \frac{r^3}{(r^4 + 4\beta^2 r^2 + 2Rr^2\cos 2\theta + R^2)^2} \,. \tag{6.62}$$

Note that f depends only on R and not on ϕ. The angular integration in (6.62) is easily performed using standard tables. Introducing the variable y, defined by $y = 4\beta^2/R$, one finds that

$$f(R) = \frac{\pi y H(y)}{16\beta^4} \,, \tag{6.63}$$

where

$$H(y) \equiv -y\frac{d}{dy}\left(\int_0^\infty \frac{dx}{\left((x^2 + yx + 1)^2 - 4x^2\right)^{1/2}} \right) . \tag{6.64}$$

All the physical details of the evolution are encoded into the function $H(y)$ and so it is worth making a few remarks about its evaluation. The denominator of the integrand in (6.64) is the square root of a quartic in x, namely

$$q(x) \equiv (x^2 + yx + 1)^2 - 4x^2 \,, \tag{6.65}$$

and so using standard techniques,[59] $H(y)$ may ultimately be expressed in terms of elliptic integrals of the first and second kinds. When $y < 4$, $q(x) = 0$ has two real roots and two complex ones; when $y > 4$ it has four real roots. Correspondingly, there are two separate calculations to perform. After some lengthy algebra, involving use of the relations between elliptic integrals and their derivatives with respect to the modulus, one finds the following.

$$\text{For } y < 4\,, \qquad H(y) = \frac{16G_1 - y^2F_1}{16 - y^2} - \frac{2}{y+4} \,, \tag{6.66}$$

$$\text{and for } y > 4\,, \qquad H(y) = \frac{4y}{16 - y^2}\,(K_2 - E_2 - F_2 + G_2) - \frac{2}{y+4}\,, \tag{6.67}$$

where

$$\begin{aligned}
F_1 &= F(k^2 = 1 - \tfrac{y^2}{16}\,,\ \psi = \tan^{-1}\tfrac{2}{\sqrt{y}})\,,\\
G_1 &= G(k^2 = 1 - \tfrac{y^2}{16}\,,\ \psi = \tan^{-1}\tfrac{2}{\sqrt{y}})\,,\\
K_2 &= K(k^2 = 1 - \tfrac{16}{y^2})\,,\\
E_2 &= E(k^2 = 1 - \tfrac{16}{y^2})\,,\\
F_2 &= F(k^2 = 1 - \tfrac{16}{y^2}\,,\ \psi = \tan^{-1}\tfrac{\sqrt{y}}{2})\,,\\
G_2 &= G(k^2 = 1 - \tfrac{16}{y^2}\,,\ \psi = \tan^{-1}\tfrac{\sqrt{y}}{2})\,,
\end{aligned} \tag{6.68}$$

with F and G the incomplete elliptic integrals of the first and second kind, respectively, defined by

$$\begin{aligned}
F(k,\psi) &= \int_0^{\psi} (1 - k^2 \sin^2\theta)^{-1/2}\, d\theta\,,\\
G(k,\psi) &= \int_0^{\psi} (1 - k^2 \sin^2\theta)^{1/2}\, d\theta\,,
\end{aligned} \tag{6.69}$$

and, as before, $K(k)$ and $E(k)$ the corresponding complete integrals obtained by taking $\psi = \pi/2$ in (6.69). The limits of (6.66) and (6.67) as $y \to 4$ from below and above are equal, namely $H(4) = \pi/8$.

Note that all the elliptic integrals in (6.68) have moduli k^2 in the range $[0,1]$, which is the conventional way of arranging things. However, if one allows k^2 to take negative values as well then it is not difficult to show that (6.66) and (6.67) are in fact equal as functions of y. The integrals K, E, F and G remain real and well defined for negative k^2, and there are no problems in their numerical evaluation. The upshot is that one can use either (6.66) or (6.67) over the *whole range* of values of y. It is natural to use (6.66), since it is slightly the simpler expression. Using NAG routines to compute G_1 and F_1, $H(y)$ is plotted in figure 6.6. It is always positive, and monotonically decreasing, with a maximum value of $\frac{1}{2}$ at $y = 0$. A little straightforward algebra reveals the behaviour at very small and very large values of y.

$$\begin{aligned}
&\text{As } y \to 0\,, \qquad H(y) = \tfrac{1}{2} + \tfrac{1}{64}\, y^2 \ln y + \mathcal{O}(y^2)\,.\\
&\text{As } y \to \infty\,, \qquad H(y) = \tfrac{2}{y}\,(\ln y - 1) + \mathcal{O}(y^{-2})\,.
\end{aligned} \tag{6.70}$$

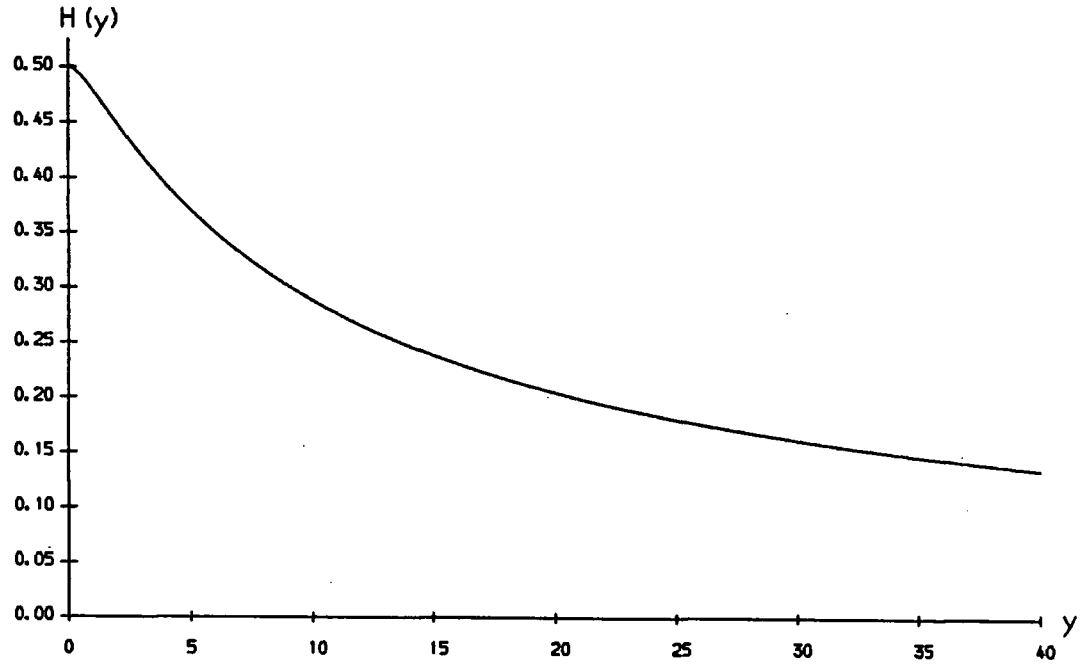


FIG 6.6. The function $H(y)$.

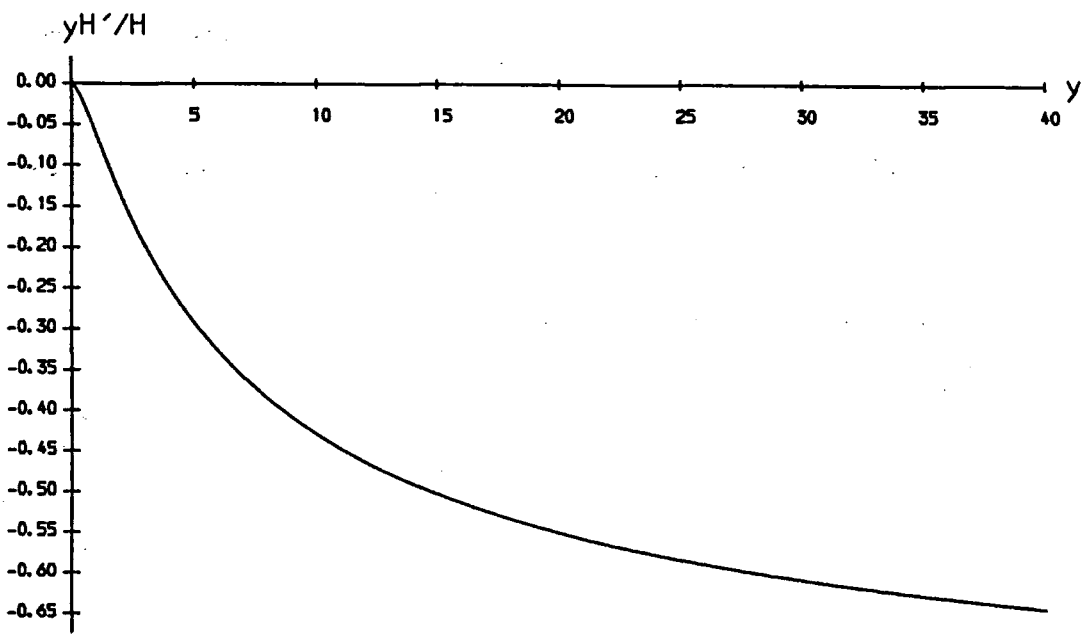


FIG 6.7. The graph of $yH'(y)/H(y)$ against y.

We are now in a position to piece together a physical description of the evolution. The natural metric associated with (6.61) may be used to obtain a picture of Σ as a surface of revolution embedded in three-dimensional Euclidean space, along the lines of chapters 12–14 of reference [56]. Beginning with a few general comments, suppose that one has a metric

$$ds^2 = a(y)^2 dy^2 + b(y)^2 d\phi^2 , \tag{6.71}$$

where a and b are positive, real-valued functions of y. Since $\partial/\partial\phi$ is a Killing vector of (6.71), it is natural to look for an embedding in which ϕ is the angle of revolution. Using cylindrical polar coordinates (ρ, ϕ, z), one wants to choose ρ and z so as to recast (6.71) in the form

$$ds^2 = d\rho^2 + dz^2 + \rho^2 d\phi^2 . \tag{6.72}$$

It is important to find any restrictions on $a(y)$ and $b(y)$ for this to be possible. Comparing (6.71) and (6.72), one must take $\rho = b$ and then require that

$$a(y)^2 dy^2 = db^2 + dz^2 . \tag{6.73}$$

The condition for z to be well defined by (6.73) is

$$a(y)^2 - b'(y)^2 \geq 0 . \tag{6.74}$$

Returning to the original problem, the natural metric coming from (6.61) is

$$ds^2 = 4\beta^2 f(R)(dR^2 + R^2 d\phi^2) . \tag{6.75}$$

Using (6.63), this may be recast in the form (6.71), with

$$a(y)^2 = \frac{4\pi\beta^2 H}{y^3} \quad \text{and} \quad b(y)^2 = \frac{4\pi\beta^2 H}{y} , \tag{6.76}$$

and then the condition (6.74) becomes

$$-1 \leq \frac{yH'(y)}{H(y)} \leq 3 . \tag{6.77}$$

The combination yH'/H may be obtained explicitly by taking $H(y)$ from (6.66) and again using standard expressions for the derivatives of the various elliptic integrals. A

short calculation yields

$$\frac{yH'(y)}{H(y)} = \frac{y^2}{16-y^2}\left(2 + \frac{(y^2-16)F_1}{16G_1 - y^2F_1 - 2(4-y)}\right). \tag{6.78}$$

This function is plotted in figure 6.7. It has a maximum value of 0, at $y = 0$, and decreases monotonically, approaching -1 as $y \to \infty$. In particular, (6.77) is satisfied, so confirming the claim that Σ may be pictured as a surface of revolution.

Now suppose that Σ is swept out by some curve C as it rotates around the z-axis. To obtain a clearer picture, it is convenient to introduce the arc length ξ along C, given by

$$d\xi^2 = d\rho^2 + dz^2 = a(y)^2 dy^2 . \tag{6.79}$$

Here, y is essentially being used as a parameter for C, whose coordinates may be thought of as $z(y)$ and $\rho(y)$. We already know that $\rho(y) = b(y)$, and so from (6.79), together with the expressions (6.76) for a and b,

$$\frac{d\rho}{d\xi} = \frac{1}{2} - \frac{yH'(y)}{2H(y)} . \tag{6.80}$$

As the parameter y varies from 0 to ∞, $d\rho/d\xi$ varies monotonically from $\frac{1}{2}$ to 1. Hence Σ looks like a round-nosed cone. The vertex of the cone, where $\rho = 0$, corresponds to $y = \infty$, or equivalently $R = 0$. Asymptotically, as $y \to 0$ and $R \to \infty$, Σ looks like a cone with angle $\pi/3$. For the sake of completeness, the curve C is drawn in figure 6.8, taking (with no loss of generality) $\beta^2 = \frac{1}{4\pi}$; in terms of the parameter y, its coordinates are given by $\rho(y) = b(y)$ and, from integrating (6.73),

$$z(y) = \int_y^\infty \sqrt{a(u)^2 - b'(u)^2}\, du . \tag{6.81}$$

It is interesting to note that round-nosed cones very similar to Σ are found in the scatterings of both monopoles[56] and vortices.[52] However, there is one qualitative difference. Recalling (6.60), we see that the vertex of Σ corresponds to $\epsilon = 0$, which as remarked earlier lies on the boundary of M_2. So strictly speaking, Σ is a round-nosed cone, punctured at the vertex.

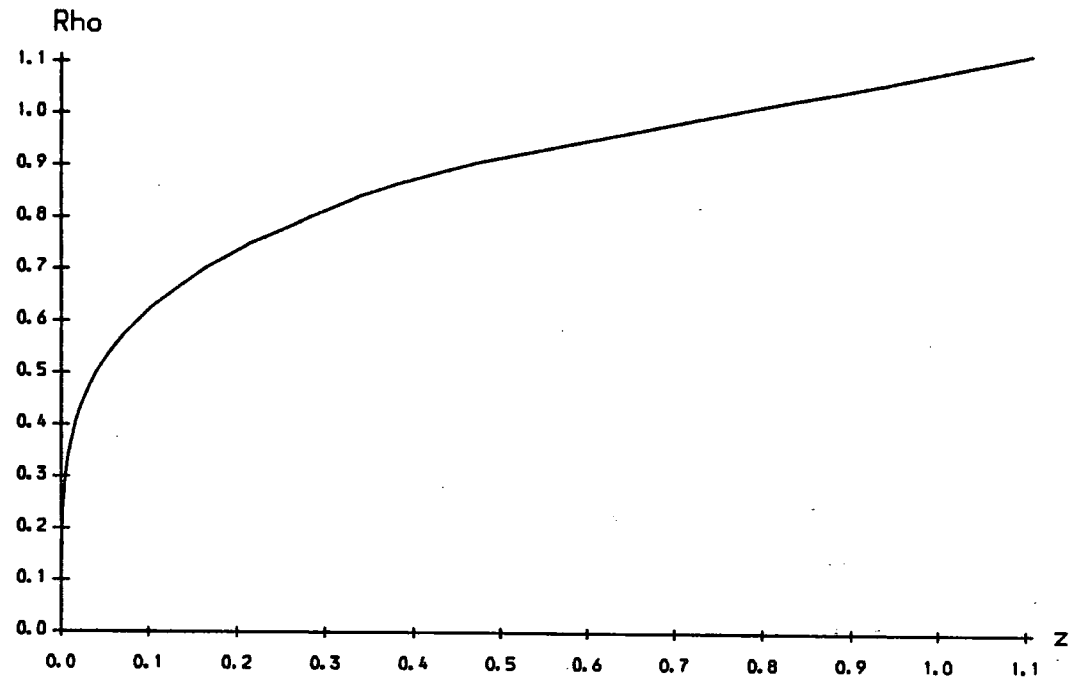


FIG 6.8. The curve C, which sweeps out the surface Σ when rotated around the z-axis.

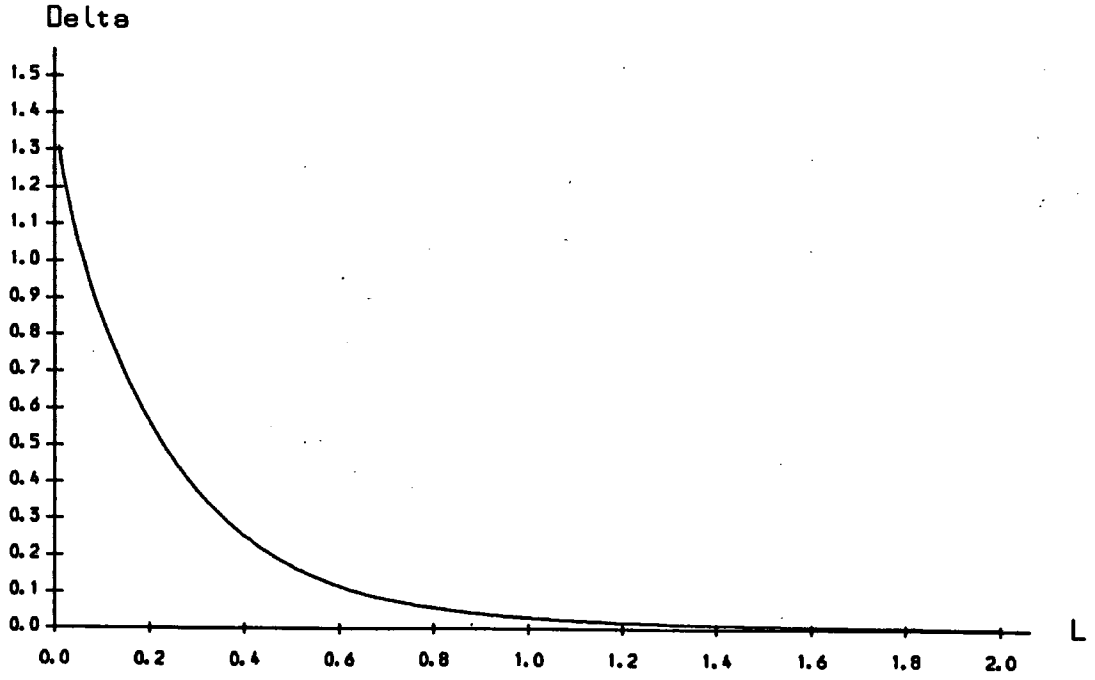


FIG 6.9. The variation of scattering angle Δ with impact parameter L.

We now have to identify the geodesics on Σ. To begin with, there are geodesics through the vertex, lying in planes that also contain the axis of Σ. As one follows such a geodesic in from the asymptotic region to the vertex, ϕ does not change; this corresponds to two solitons in a head-on collision. At the vertex, the solitons collapse to infinitely tall spikes. However, if the vertex were present then the geodesic would emerge with ϕ having changed by π. Recall that (asymptotically, at least) the soliton positions are $\pm i\epsilon^{1/2}$, so one would have scattering at right angles. Although the vertex of Σ is missing, one expects there to be geodesics close to these, corresponding to small impact parameters, with scattering angles close to 90°.

Unlike the situation described above, the general geodesic does not lie in a plane, but winds around Σ, *i.e.* it does not have a constant value of ϕ. The scattering angle, Δ (positive for repulsive interactions and negative for attractive ones), is related to the change in ϕ by

$$\Delta = \pi - \frac{1}{2}\int_0^\infty \dot{\phi}(t)\,dt\,. \tag{6.82}$$

To get a handle on (6.82) in terms of physical parameters, observe that the Killing vector $\partial/\partial\phi$ of the metric (6.75) generates a second constant of motion, L say, in addition to the kinetic energy T. Explicitly,

$$\begin{aligned} L &= R^2 f(R)\dot{\phi}\,, \\ T &= 4\beta^2 f(R)\,(\dot{R}^2 + R^2\dot{\phi}^2)\,. \end{aligned} \tag{6.83}$$

We shall see shortly that L is essentially the impact parameter of the collision, and can be used to label geodesics on Σ. Without loss of generality one can fix the initial soliton velocities so that $T = 4\beta^2/\pi$. This does not affect the scattering angle, only the rate at which the geodesics are traced out. Then eliminating $\dot{\phi}$ from (6.83), and using (6.63) to replace R with y, one finds

$$\dot{y}^2 = \frac{y^3}{\pi^2 H(y)^2}\left(H(y) - L^2 y\right). \tag{6.84}$$

From figure 6.6 it is clear that for positive values of L, $H - L^2 y = 0$ has a single positive root, at $y = y_p$ say. When $y = y_p$, $\dot{y} = 0$. As a geodesic comes in from the asymptotic region, y increases from 0 until it reaches a maximum value of y_p at some time

$t = t_p$. The geodesic then returns to infinity, its two halves being mirror images of each other. Suppose that at $t = 0$ the solitons are essentially at infinity (in the language of section 6.3 this means taking $\lambda \to \infty$). Then equation (6.82) may be rewritten

$$\Delta = \pi - \int_{t_p}^{\infty} \dot{\phi}(t)\, dt\,. \tag{6.85}$$

Using (6.83) and (6.84), the integrand may be written entirely in terms of y, leaving Δ as a function of L:

$$\Delta(L) = \pi - L \int_0^{y_p} \frac{dy}{\sqrt{y}\sqrt{H(y) - L^2 y}}\,. \tag{6.86}$$

The results of this section are summarized in figure 6.9, where Δ is plotted against L. This curve is derived from (6.86) using NAG routines, both to find y_p and then to calculate the integral. The first thing to note is that Δ is always positive, *i.e.* the interaction is always repulsive. Second, as argued earlier, Δ approaches $\pi/2$ at small impact parameters. This type of variation of scattering angle with impact parameter is familiar from studies of vortices and monopoles. Compare figure 6.9 with, for example, figure 3 of reference [52] and diagram 16 of reference [56].

It only remains to clarify the interpretation of L as the impact parameter. To do this we shall calculate L from the initial data, which are denoted (as in section 6.4) by a subscript 0. Comparing (6.60) with (6.19), then using (6.23) and (6.30), shows that

$$R_0 = \lambda^2 w^2 \quad \text{and} \quad \dot{\phi}_0 = \frac{2vJ}{\lambda^2 w}\,. \tag{6.87}$$

Taking $\lambda \to \infty$, we see that $R_0 \to \infty$ and so, from (6.63) and (6.70), $f(R_0) \to \pi/(8\beta^2 R_0)$. Substituting into (6.83) gives

$$L = \frac{J\pi v w}{4\beta^2}\,, \tag{6.88}$$

and so it is natural to interpret L as an angular momentum. In particular, if the initial values of the width and velocity are kept fixed then L is simply a constant multiple of the impact parameter, J.

6.6 Numerical Results : Two-Dimensional Evolutions

There are some two-dimensional geodesics which are not covered by sections 6.4 and 6.5. They arise from evolving initial data of the form (6.39), (6.40) or (6.41), when β is nonzero. To illustrate some of the possibilities we shall take one example from each case. First, consider initial data given by (6.39), with

$$\beta = \delta = 0.3\,, \quad w = 1.0\,, \quad \lambda = 5.0\,, \quad d_+ = 0\,, \quad v = 0.1\,. \tag{6.89}$$

The results of evolving (6.89) are shown in figure 6.10. Initially, the solitons have widths 0.7 and 1.3. After scattering, they emerge at right angles to the original motion, and with the same width. That this should happen is clear from (6.21), since $\epsilon^{1/2}$ evolves from being pure imaginary to pure real. What is not predicted analytically is the subsequent motion. As they emerge from the interaction, the solitons shrink, reach a minimum width, and then expand, apparently without limit. Further numerical experiments show that the minimum width attained is roughly proportional to the difference in their initial widths. The limiting case of this is, of course, figure 6.4, in which the solitons become spikes, *i.e.* they attain zero width.

Next, we shall keep all the physical parameters of (6.89), but use a different value of β, so that the initial data are given by (6.40), with

$$\beta = w = 1.0\,, \quad \delta = 0.3\,, \quad \lambda = 5.0\,, \quad d_+ = 0\,, \quad v = 0.1\,. \tag{6.90}$$

The evolution is now very different (figure 6.11). A singularity develops in the energy density as the solitons approach each other. In figure 6.11, this is shown by the sharp decrease in width of the smaller one as it comes close to the origin. It is not clear from the picture, but it may be that the singularity occurs at the origin, when the two soliton positions coincide. Note that all pairs of circles are drawn, regardless of whether they intersect, so that one can see the singularity developing.

Finally in this section, consider evolving initial data of the form (6.41), with

$$\beta = 0.6\,, \quad w = 1.0\,, \quad \lambda = 5.0\,, \quad d_+ = 0\,, \quad v = 0.1\,. \tag{6.91}$$

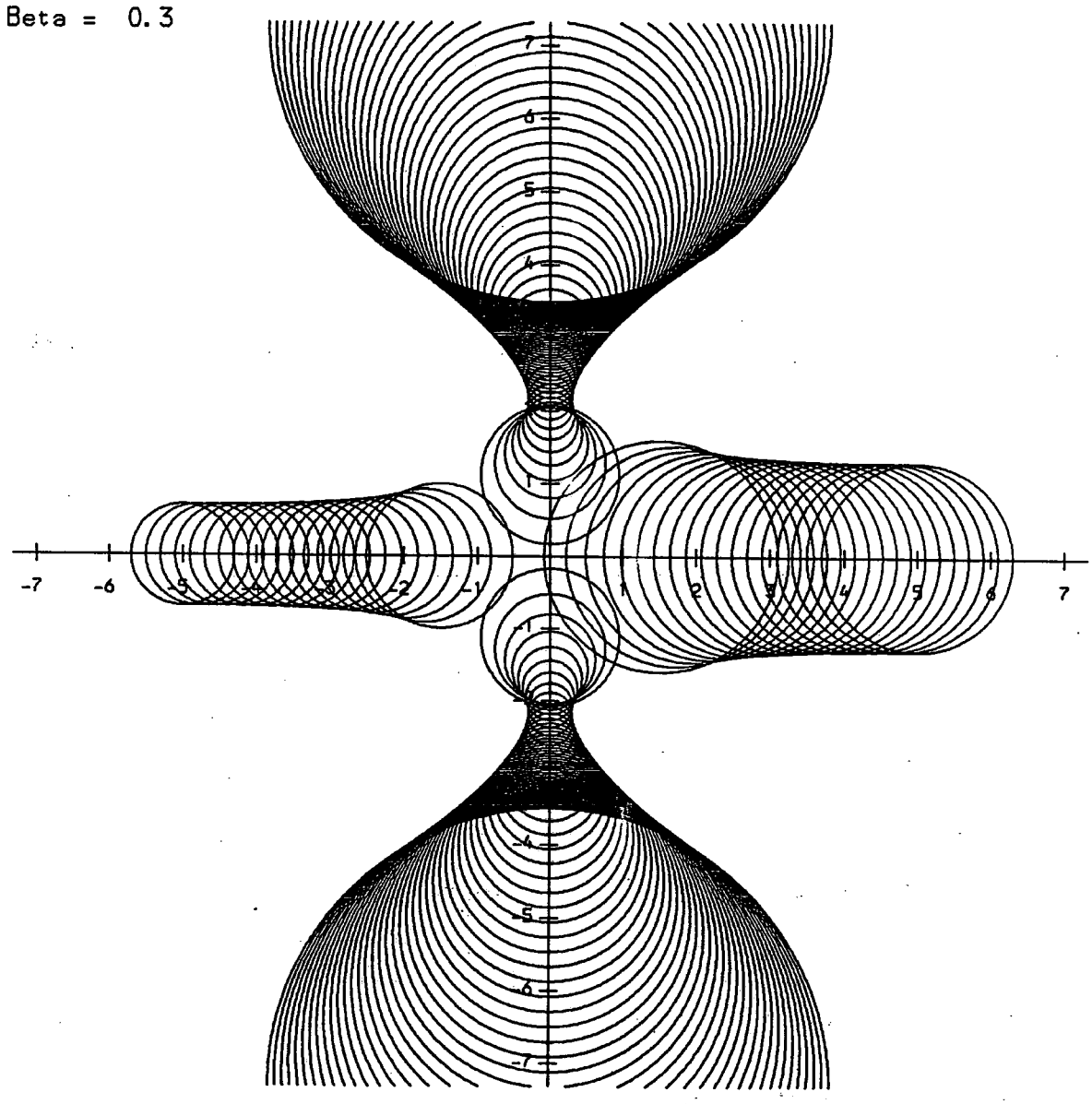


FIG 6.10. Evolution of the initial data (6.89).

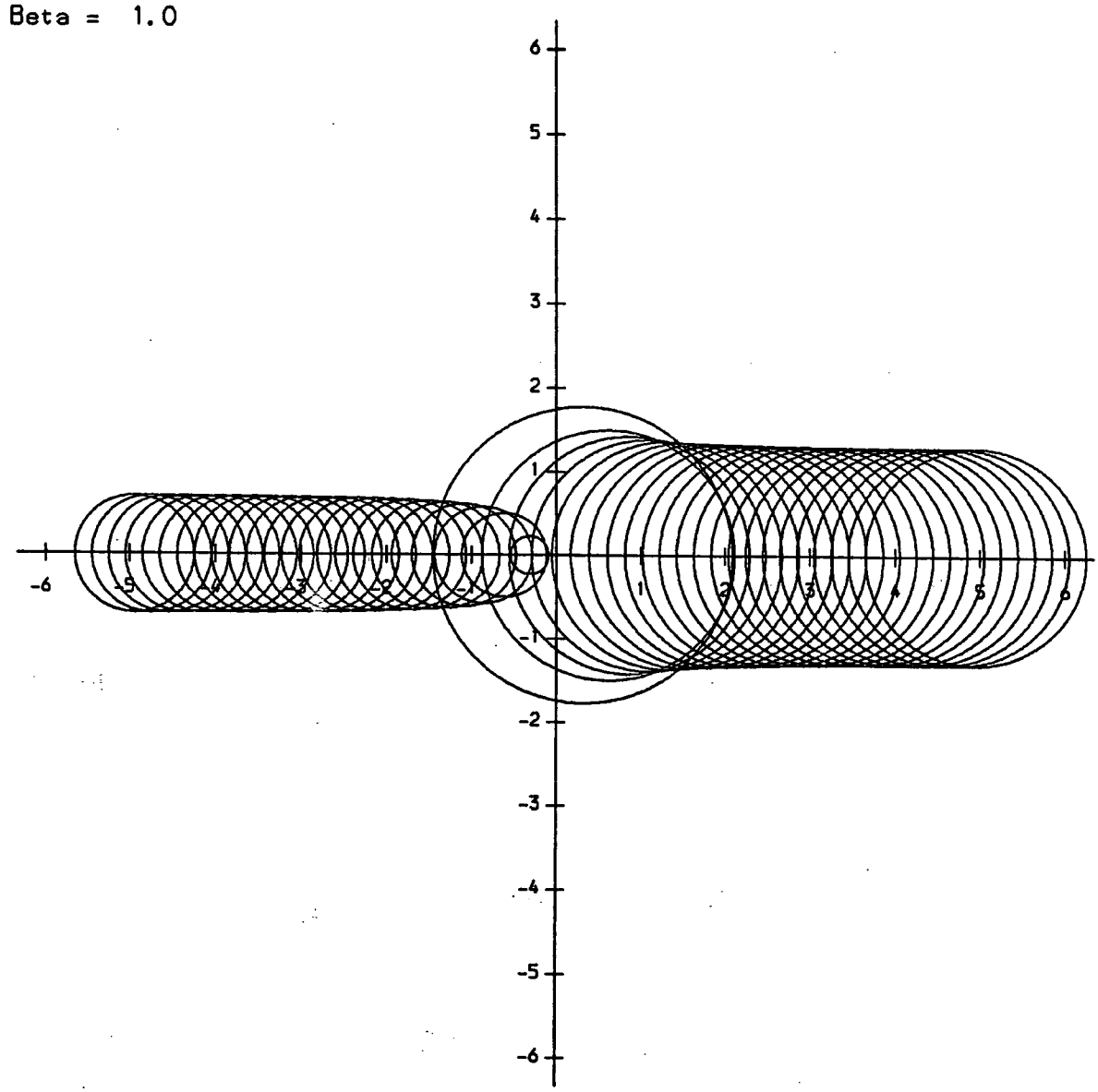


FIG 6.11. Evolution of the initial data (6.90).

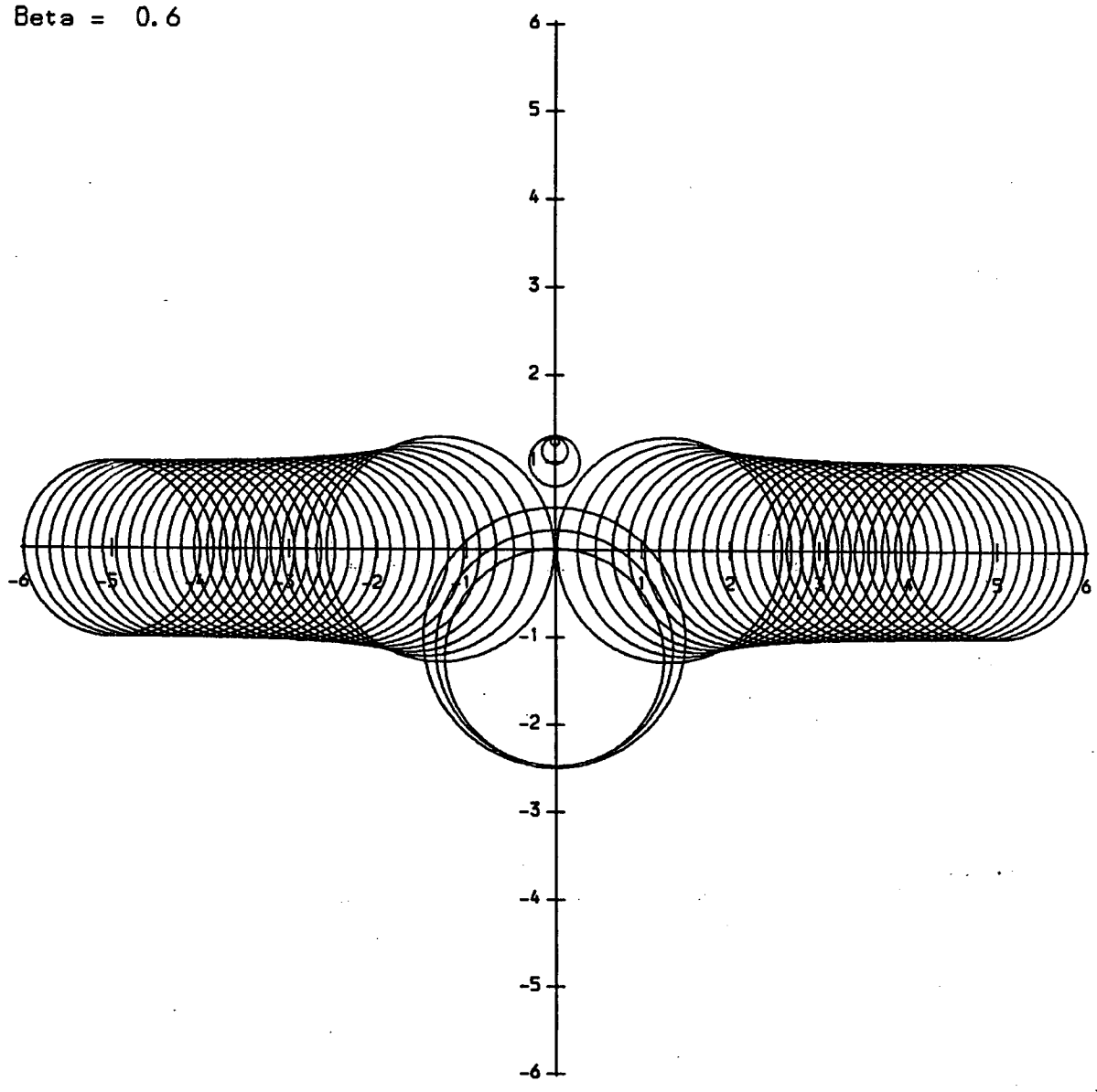


FIG 6.12. Evolution of the initial data (6.91).

The results of this calculation are shown in figure 6.12. Here, the solitons have the same width initially, but emerge from the interaction with *different* sizes (as predicted by (6.21), since γ is now pure imaginary). The smaller one then rapidly shrinks to a spike. In many ways, the evolution is similar to that of figure 6.4, in which the solitons shrink to spikes simultaneously.

6.7 Numerical Results : Four-Dimensional Evolutions

In this section, we move away from the various two-dimensional submanifolds and look at some more general geodesics on M_2. However, we shall restrict attention to case (i) of section 6.3 (*i.e.* $\beta = \delta = 0$ and $d_+ = d_-$), for which it has been shown by Ward[46] that there are two constants of motion in addition to the kinetic energy. Explicitly, if one writes

$$\begin{aligned} \gamma &= R\sin\psi e^{i\sigma}\,, \\ \epsilon &= R\cos\psi e^{i\phi}\,, \end{aligned} \tag{6.92}$$

then $\partial/\partial\sigma$ and $\partial/\partial\phi$ are both Killing vectors of the natural metric, and the corresponding constants of motion are

$$\begin{aligned} L &= 2R\omega\dot{\phi} + R\mu\dot{\sigma}\,, \\ M &= 2R\nu\dot{\sigma} + R\mu\dot{\phi}\,, \end{aligned} \tag{6.93}$$

where

$$\begin{aligned} \omega &\equiv \tfrac{1}{2}\cos^2\psi\, E(\cos\psi)\,, \\ \mu &\equiv -\sin^2\psi\left(K(\cos\psi) - E(\cos\psi)\right), \\ \nu &\equiv \sin^2\psi\left(K(\cos\psi) - \tfrac{1}{2}E(\cos\psi)\right), \end{aligned} \tag{6.94}$$

with K and E the complete elliptic integrals of the first and second kinds, as before.

Throughout section 6.4, L and M were both zero, so it is now natural to look at the consequences of taking nonzero values. There is an intuitive interpretation in terms of angular momentum, in which one thinks of a spatial angular momentum, $\mathcal{J}_S$ (corresponding to nonzero impact parameters), and an "internal" angular momentum, $\mathcal{J}_I$ (which takes account of the time-dependence of the phase σ). Note from (6.35) that the initial value of $\dot{\sigma}$ (and hence that of $\mathcal{J}_I$) may be specified independently of the physical parameters.

The precise definitions of $\mathcal{J}_S$ and $\mathcal{J}_I$ are not immediately clear, but this does not rule out a heuristic discussion. From the asymptotic picture (6.21), one expects that

$$\begin{aligned} \mathcal{J}_S &= Rf(\psi)\dot{\phi}\,, \\ \mathcal{J}_I &= Rg(\psi)(2\dot{\sigma} - \dot{\phi})\,, \end{aligned} \tag{6.95}$$

where f and g are some functions of ψ alone. Then (6.93) may be written

$$\begin{aligned} L &= p(\psi)\mathcal{J}_S + q(\psi)\mathcal{J}_I\,, \\ M &= r(\psi)\mathcal{J}_S + s(\psi)\mathcal{J}_I\,, \end{aligned} \tag{6.96}$$

for some p, q, r and s. The general idea is that p, q, r and s vary with time, because ψ does; $\mathcal{J}_S$ and $\mathcal{J}_I$ are conserved in the linear combinations (6.96), but not individually. In other words, there is an exchange between the spatial and internal angular momenta. This type of behaviour is familiar from the scattering of magnetic monopoles, in particular from the type II scattering mentioned in section 6.1.

In the first example, $\mathcal{J}_I$ is equal to zero initially, but $\mathcal{J}_S$ is nonzero. Using the notation of section 6.3, the initial data correspond to the physical parameters

$$w = 1.0\,, \quad J = 0.5\,, \quad \lambda = 10.0\,, \quad d = 0\,, \quad v = 0.1\,, \quad \dot{\sigma} = 0\,, \tag{6.97}$$

where d is the initial dilation of *each* soliton (recall that $d_+ = d_-$). The evolution is shown in figure 6.13. The solitons scatter at some positive angle, but then shrink to spikes a short time later. In this respect, things are very similar to figure 6.4.

Next, consider $\mathcal{J}_S$ initially zero but $\mathcal{J}_I$ nonzero. This corresponds to a head-on collision, but with the phase σ given an initial time dependence. Figure 6.14 shows the results of taking $\dot{\sigma} = 0.02$, together with the same physical parameters as figure 6.4, namely

$$w = 1.0\,, \quad J = 0\,, \quad \lambda = 5.0\,, \quad d = 0\,, \quad v = 0.1\,. \tag{6.98}$$

This interaction seems quite unlike any occurring in other soliton theories. The scattering angle is approximately 106°, *i.e.* the solitons "bounce back" beyond the perpendicular. Smaller initial values of $\dot{\sigma}$ give scattering angles closer to 90°, while larger values cause the solitons to bounce back even more. In the central region, $\mathcal{J}_S$ becomes nonzero, as the solitons move off the x-axis. However, asymptotically they appear to have emerged from the origin, *i.e.* $\mathcal{J}_S$ vanishes as $t \to \infty$. This observation gives rise to the conjecture that, although $\mathcal{J}_S$ and $\mathcal{J}_I$ are not conserved throughout, they *are* conserved asymptotically.

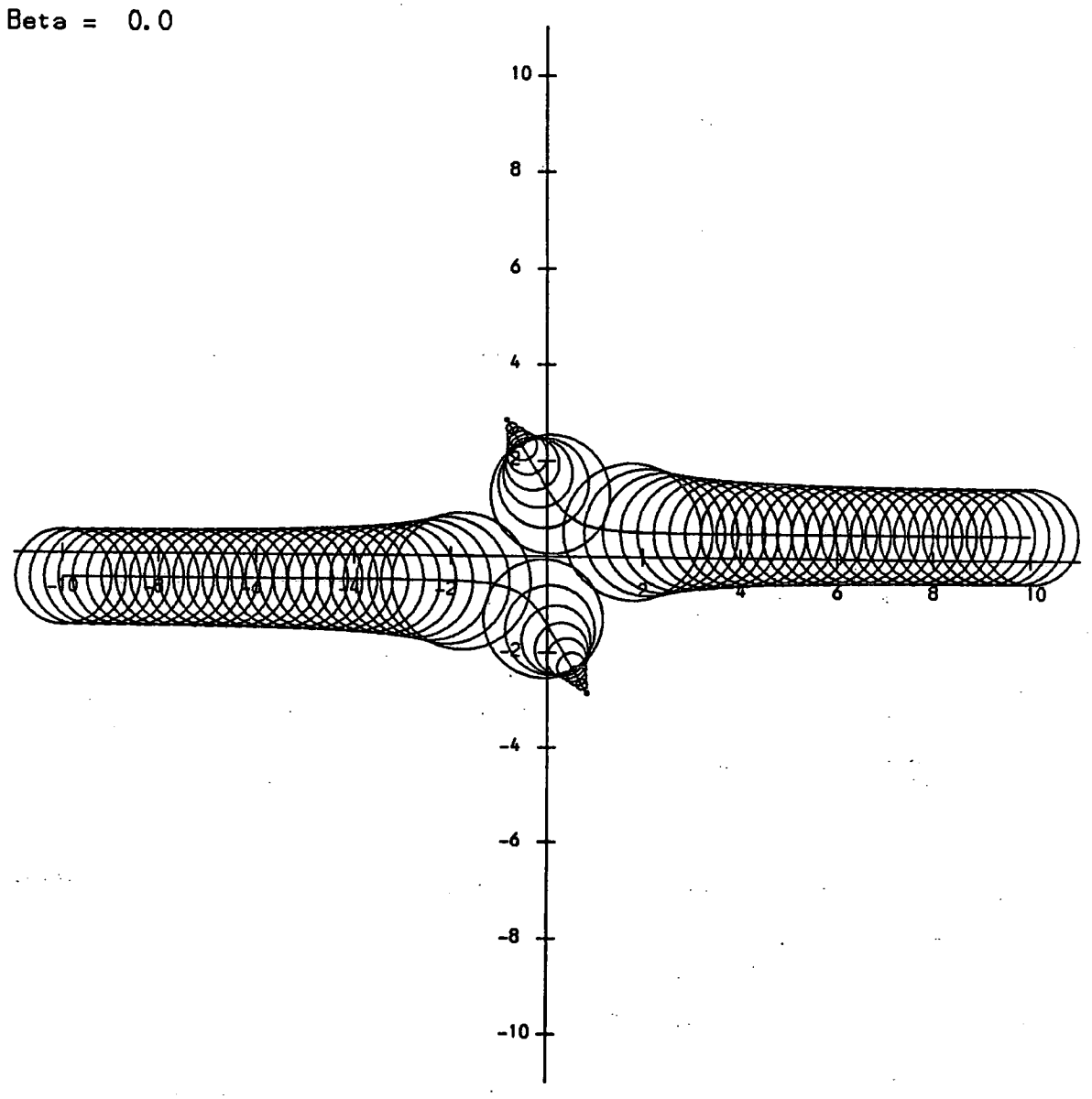


FIG 6.13. Evolution of the initial data (6.97).

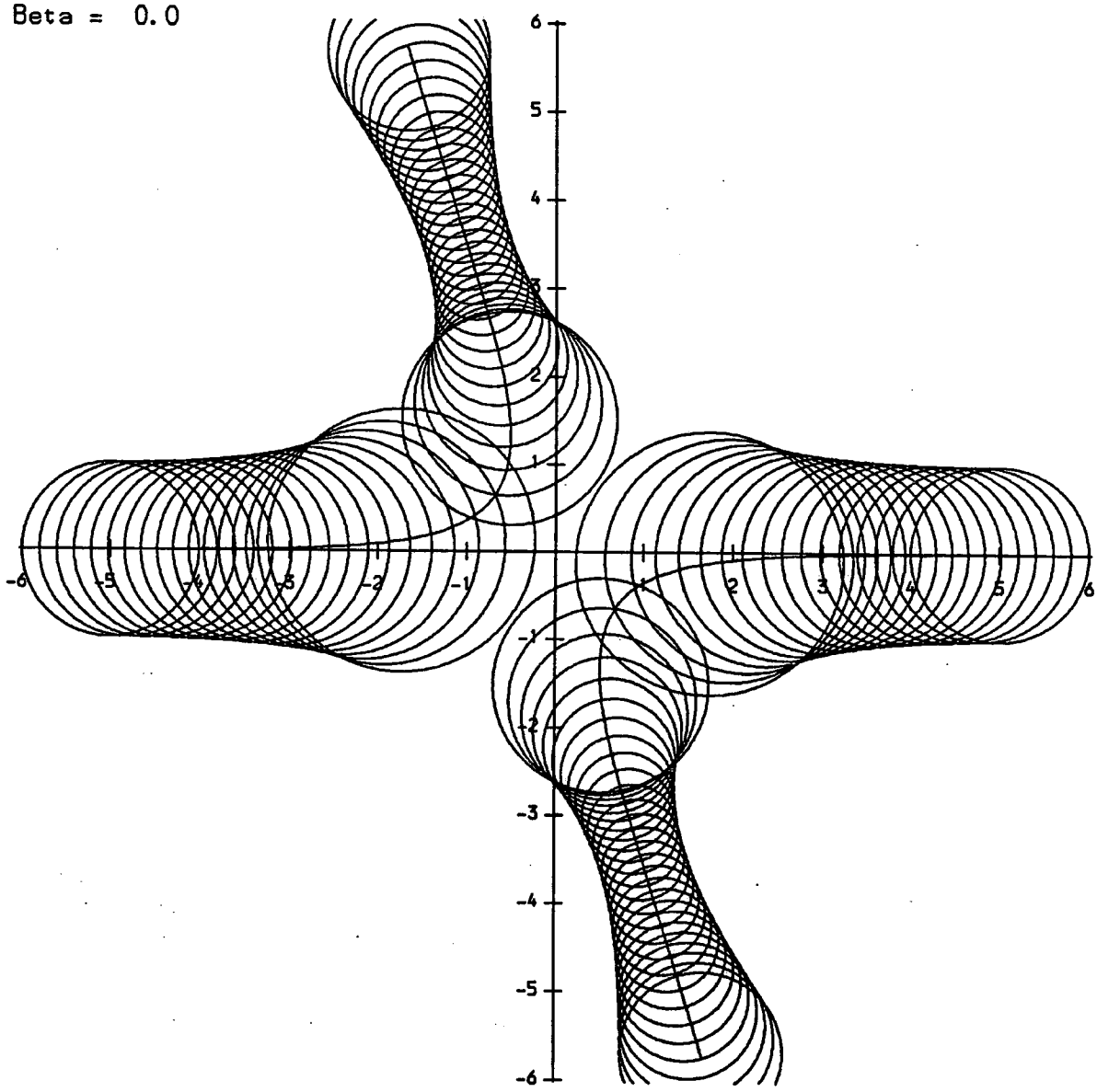


FIG 6.14. Evolution of the initial data (6.98).

Finally, we consider evolutions in which $\mathcal{J}_S$ and $\mathcal{J}_I$ are both nonzero initially. In fact, some geodesics of this type were found analytically by Ward,[46] although at the time it was not known how they fitted into the general scheme of things. They correspond to a constant value of ψ in (6.92), with γ and ϵ of the form

$$\begin{aligned} \gamma(t) &= a \sin\psi \big(b + i(t-\alpha)\big)^2 , \\ \epsilon(t) &= a \cos\psi \big(b + i(t-\alpha)\big)^2 , \end{aligned} \tag{6.99}$$

for some real constants a, b and α. Since $\epsilon^{1/2}$ is linear in t, the solitons move in straight lines at constant speed, a behaviour more usually associated with integrable models. The soliton width is given by

$$w = \tfrac{1}{2}\Big(a \sin\psi \tan\psi \big(b^2 + (t-\alpha)^2\big)\Big)^{1/2} , \tag{6.100}$$

i.e. the solitons shrink as they approach, reaching a minimum width at $t = \alpha$ (the point of closest approach), after which they expand as they move apart. Moreover, since ψ is constant, (6.96) shows that in these special cases $\mathcal{J}_S$ and $\mathcal{J}_I$ are conserved individually. Figure 6.15 illustrates one particular process, obtained by evolving the initial data

$$w = 2.0\,, \quad J = 1.5\,, \quad \lambda = 2.5\,, \quad d = -0.064\,, \quad v = 0.2\,, \quad \dot{\sigma} = 0.048\,, \tag{6.101}$$

which correspond to

$$\begin{aligned} \gamma(t) &= \tfrac{4}{125}\big(15 + i(t-20)\big)^2 , \\ \epsilon(t) &= \tfrac{1}{25}\big(15 + i(t-20)\big)^2 . \end{aligned} \tag{6.102}$$

That this process should be reproduced so accurately is a further indication that the numerical implementation is sound.

As a comparison, figure 6.16 shows the results of evolving the same initial data, but now taking $d = 0$. The solitons now experience a deflection as they move by each other. So it is seen that a carefully chosen initial squashing, as in (6.101), can have the effect of "absorbing" the forces of repulsion that would otherwise be present.

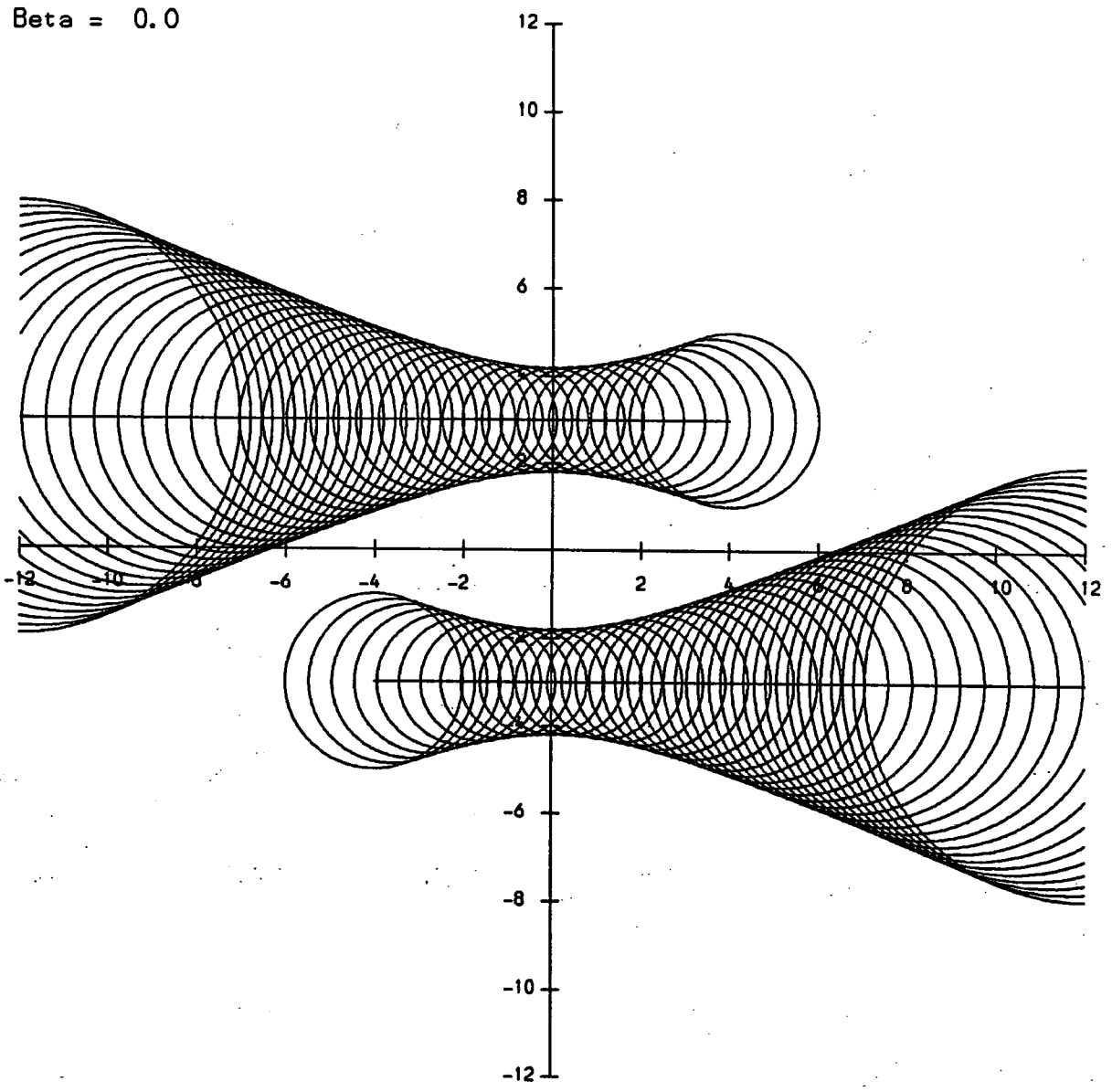


FIG 6.15. Evolution of initial data (6.101).

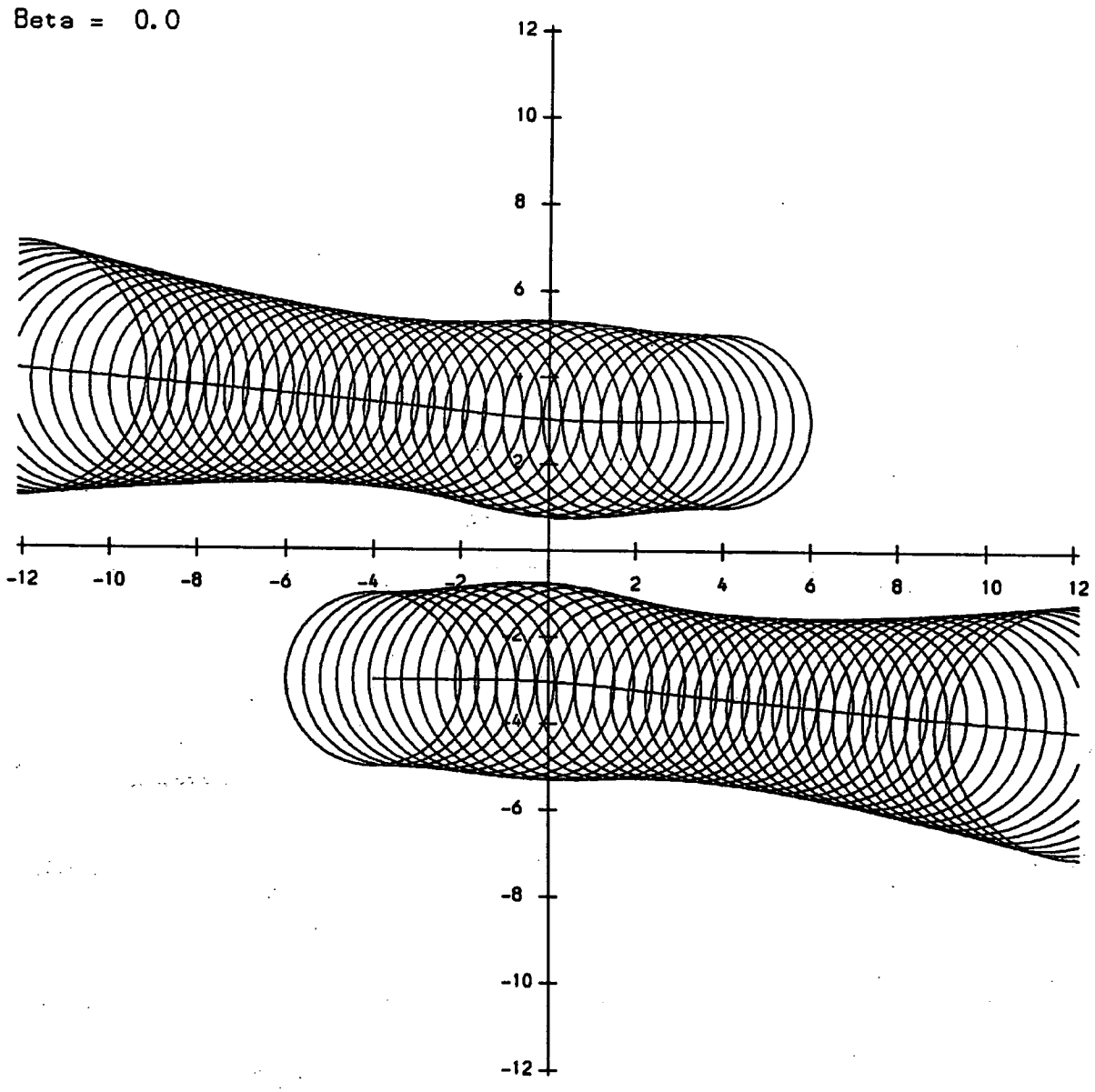


FIG 6.16. Evolution of initial data (6.101) with d now set to zero.

6.8 Concluding Remarks

To sum up, we have classified all possible sets of initial data in terms of physical quantities, identifying in addition the various "internal" parameters, which do not change the initial energy density, but which are important in understanding the evolution as a whole. Using the geodesic approximation, some sets of initial data were investigated analytically. Other, more exotic, possibilities were studied numerically. It is clear that the $\mathbb{C}P^1$ model exhibits scattering processes quite unlike those found in other soliton theories. The wide diversity of behaviour stems from the conformal invariance of the model, which allows solitons to change size as they move. In view of the current interest of the $\mathbb{C}P^1$ model in relation to high temperature superconductivity[10] and the quantum Hall effect,[11] there is also the exciting possibility that one might even be able to find experimental evidence for some of these interactions.

In principle, the computer codes that have been developed could be used to compile a library of all possible scattering processes, but this would be a formidable undertaking. It seems better to try to explain the apparent complexity by finding a few general rules. For example, it might be the case that the interaction is always repulsive. Alternatively, perhaps one could find a condition for the solitons to remain of finite width, *i.e.* for geodesics to avoid the boundary of M_2. Since M_2 is four-dimensional and its boundary two-dimensional, one expects that the formation of spiky configurations will be the exception rather than the rule. The results presented here support the conjecture that singularities in the energy density will not develop when $\dot{\sigma}$ is initially nonzero, in keeping with the findings of chapter V, but at the moment no proof exists. It seems that answers to these questions will only come with a better understanding of the geometry of M_2.

One should also bear in mind that the evolutions presented here were obtained using an approximation, the validity of which has never been rigorously established. Experience with vortices and monopoles suggests that radiative effects will be small for a wide range of initial velocities and dilations, but this should nevertheless be looked at in a numerical simulation of the full $\mathbb{C}P^1$ model, such as that recently carried out by Zakrzewski, for which detailed results will be reported soon.[60] In numerical studies of this type, the fields are evolved on a two-dimensional lattice using a discrete analogue of the full field equations. Briefly, Zakrzewski's work shows, as hoped, that the amounts of radiation emitted during scattering processes are negligible. Unfortunately, a quantitative comparison with

the geodesic approximation is hampered by subtle numerical problems. It appears that there are always small perturbations, arising from the lattice, which can build up over the course of an evolution to give a large effect. For example, in Zakrzewski's implementation, the solitons always shrink to zero width after interacting. Nevertheless, there is still qualitative evidence supporting all the processes discussed here, with the exception of that shown in figure 6.14, which remains the subject of some debate.

As far as suggestions for future work are concerned, two projects come to mind. The first is to see how things get modified by quantum effects, perhaps using techniques similar to those described for monopoles in reference [61].

The second has already been mentioned at the end of chapter V, namely the possibility of modifying the model, in an attempt to prevent singularities ever developing in the energy density. Further comment is deferred until the end of chapter VII.

Chapter VII
Some Open Questions

At first sight the integrable chiral model of chapter III is very different from the conventional O(3) model considered in the last three chapters; but in fact they are very closely related. For example, they possess essentially the same static solitons. In the formalism of the integrable model the static solutions come from taking $\mu = i$ and correspond to a matrix J of the form

$$J = \frac{i}{1+|f|^2}\begin{pmatrix} 1-|f|^2 & 2f \\ 2\bar{f} & -1+|f|^2 \end{pmatrix}, \tag{7.1}$$

for some meromorphic function $f(z)$. It is easy to check that any such matrix satisfies $J^2 = -\mathbb{1}$, *i.e.* they all lie on the equator of SU(2), which is precisely the condition for reduction to the O(3) model. The similarity becomes even more explicit when one considers the expressions for the potential energy in each case. In the integrable model, equation (3.10) shows that the configurations (7.1) have potential energy

$$E[f] = \int \frac{8|f'|^2}{(1+|f|^2)^2}\, d^2x\,, \tag{7.2}$$

while for the O(3) model the corresponding expression (parametrizing the two-sphere using the complex field u) is

$$E[u] = \int \frac{|\partial_x u|^2 + |\partial_y u|^2}{(1+|u|^2)^2}\, d^2x\,.$$

It is now seen that for the time-independent solutions, f in the integrable model plays the same role as u in the O(3) model.

With the above observations in mind it is natural to ask the complementary questions to the ones answered in this thesis. How do extended waves behave in the O(3) σ-model? What are the stability properties of lumps in the integrable chiral model? The next two paragraphs outline what is known about these problems at the moment.

As far as extended waves in the O(3) σ-model are concerned, the simplest such solution is $u = \exp(z)$, which gives a wave that is static and lies along the y-axis. It can be made to move by Lorentz boosting. To get a wave along some other direction, one would take $u = \exp(bz)$, where b is a complex parameter. This is in direct analogy with the construction of waves in the integrable chiral model. However, the question of wave interactions presents some new problems. A numerical evolution of the full field equations would require a very careful choice of boundary condition. On the other hand, for a slow-motion approximation, one has the problem of identifying the full manifold of static two-wave solutions, and also the problem that waves have infinite energy. Nevertheless, some progress should still be possible.

Turning now to the question of soliton stability in the integrable chiral model, a static lump will, of course, have the same zero modes as in the O(3) model, *i.e.* it is capable of shrinking or expanding indefinitely under small perturbations. But, in addition, the absence of a topological stability suggests the existence of *negative* modes, which correspond to the field J moving off the equator of SU(2). Because the model is integrable, one might expect that these modes could be constructed explicitly. A considerable amount of work has been done looking at small perturbations of the form

$$J = J_0(\mathbb{1} + i\boldsymbol{b}\,.\,\boldsymbol{\sigma}) + \mathcal{O}(\boldsymbol{b}^2)\,, \tag{7.3}$$

where J_0 is the chiral field for a static lump, and the perturbation $\boldsymbol{b}$ is some function of x, y and t. The construction of explicit negative modes has proved to be difficult. Some results have been obtained restricting the form that $\boldsymbol{b}$ must take, but so far no negative mode has been found. Bearing in mind that scattering processes do not appear to excite any negative mode, it is conceivable that none exist. At the moment this remains an open question.

We now return in a little more detail to the questions raised at the ends of chapters V and VI, namely the possibility of modifying the O(3) model so as to stabilize the solitons,

both under the influence of small perturbations and in interactions. As mentioned in chapter V, there appear to be two natural options, which are discussed in turn in the next few paragraphs.

Firstly, one could add to the Lagrangian both a Skyrme-like term (containing higher powers of the field derivatives) and a potential term. It is easy to see from the scaling arguments discussed in chapter II that the Skyrme term drives the soliton towards a large width, wheras the potential term encourages it to shrink. It is these two competing effects that can lead to a stable configuration. As an explicit example, consider the Lagrangian

$$\mathcal{L}' = \mathcal{L}_0 + \tfrac{1}{4}\theta_1\big((\partial_\mu \boldsymbol{\phi}\,.\,\partial^\mu \boldsymbol{\phi})^2 - (\partial_\mu \boldsymbol{\phi}\,.\,\partial_\nu \boldsymbol{\phi})(\partial^\mu \boldsymbol{\phi}\,.\,\partial^\nu \boldsymbol{\phi})\big) + \tfrac{1}{4}\theta_2(1+\phi_3)^4\,, \tag{7.4}$$

where

$$\mathcal{L}_0 = \tfrac{1}{4}(\partial_\mu \boldsymbol{\phi}\,.\,\partial^\mu \boldsymbol{\phi}) \tag{7.5}$$

is the Lagrangian for the unmodified model, and θ_1 and θ_2 are (positive) real parameters. The term proportional to θ_1 is the Skyrme term, and contains fourth powers of the field derivatives. Its form is essentially unique, provided one requires that the Hamiltonian be positive definite, and that time derivatives appear only quadratically (so that the equations of motion may be calculated in the usual way). The main freedom lies in the choice of potential (*i.e.* in the term proportional to θ_2). As we shall see shortly, it has been chosen here so that there is a simple analytic solution of the field equations.

For the purposes of calculation, it is convenient to rewrite (7.4) in terms of the $\mathbb{C}\mathrm{P}^1$ field u. After a little work, one finds that

$$\begin{aligned}\mathcal{L}' = \mathcal{L}_0 + 8\theta_1(1+u\bar{u})^{-4}\Big(\big(\mathrm{Im}(\bar{u}_x u_y)\big)^2 - \big(\mathrm{Im}(\bar{u}_t u_y)\big)^2 - \big(\mathrm{Im}(\bar{u}_t u_x)\big)^2\Big) \\ + 4\theta_2(u\bar{u})^4(1+u\bar{u})^{-4}\,,\end{aligned} \tag{7.6}$$

where

$$\mathcal{L}_0 = \frac{\partial_\mu u\,\partial^\mu \bar{u}}{(1+u\bar{u})^2}\,. \tag{7.7}$$

The derivation of the field equations is straightforward, albeit a little messy. It is not difficult to check that one particular solution, with topological charge equal to one, is

$$u = \frac{\lambda}{z}\,, \qquad \text{where} \qquad \lambda = \sqrt[4]{\frac{2\theta_1}{\theta_2}}\,. \tag{7.8}$$

The most striking thing about (7.8) is that it is also a solution of the unmodified model.

This is a consequence of the careful choice of potential in (7.6). However, note that no other value of λ will do, *i.e.* the soliton width is fixed in terms of the parameters θ_1 and θ_2. Moreover, one can see that if $\theta_1 \to 0$ then $\lambda \to 0$, since the potential term now dominates. Conversely, if $\theta_2 \to 0$ then the Skyrme term dominates and $\lambda \to \infty$.

It is natural to refer to the solution (7.8) as a "skyrmion". The stability of this skyrmion to small perturbations has been demonstrated numerically,[62] but it is not known whether there are any similar solutions with higher values of the topological charge. Certainly, all attempts to construct such objects analytically have failed; and further numerical work[62] indicates the presence of forces between two isolated skyrmions, which tends to rule out the possibility of static two-lump configurations. Perhaps the best guess is that there *are* static solutions of higher topological charge, but that they are all radially symmetric and can only be constructed numerically.

It is worth mentioning that a Lagrangian very similar to (7.4) has recently appeared in a paper by Bogolubskaya and Bogolubsky.[63] The only difference is that the potential term is taken to be of the form $(1-\phi_3^2)$, a choice motivated by considering the anisotropic Heisenberg ferromagnet. In this model, even the charge-one skyrmion must be found numerically, although it is interesting to note that the two Russians also managed to calculate a radially symmetric solution of charge two.

Finally, we mention some preliminary results on the scattering of skyrmions, obtained via numerical evolution of the full field equations.[62] For a head-on collision of two skyrmions, there seems to be a critical impact velocity. Below the critical velocity, they bounce back along their original trajectories; above the critical velocity they scatter at 90°. For velocities only just below the critical value there is a short-lived bound state. These features are very reminiscent of the ϕ^4 theory in (1+1) dimensions.[64,65] Clearly this subject merits further investigation.

The second, and perhaps more elegant, modification of the O(3) model also involves the addition of a potential term, but does not include a Skyrme term. To counteract the tendency of the solitons to shrink, the fields are given an internal rotation, in such a way that the energy density remains independent of time. Objects of this kind have become known in the literature as "Q-balls", following the work of Coleman[21] and others[66]

(although these original Q-balls lived not in two, but three spatial dimensions). Their stability depends on the existence of a conserved Noether charge, resulting from an unbroken continuous global symmetry of the Lagrangian.

Strictly speaking, Q-balls arise when this is the *only* symmetry; but there may be other, spontaneously broken symmetries, and in these cases one tends to use the collective term "nontopological solitons". For example, a model in which there is a spontaneously broken discrete symmetry has been discussed recently in reference [67]. A similar model was discussed a few years ago by Friedberg, Lee and Sirlin,[68] who also went on to look at models containing gauge[69] and fermion[70] fields. More recently, the same ideas have been used in the construction of so-called "soliton stars".[71] These are cold, stable stellar configurations, which can have very large masses (much larger than neutron stars or white dwarves), without becoming black holes.

To get back to the matter in hand, the objects discussed below are slightly different; first because they live in just two spatial dimensions, second because they have both a Noether charge Q, *and* the usual topological charge k. In this sense they have both topological and nontopological aspects. To distinguish them from conventional Q-balls, they will be referred to as "Q-lumps". Taking a specific example, consider the Lagrangian

$$\mathcal{L}_Q = \mathcal{L}_0 + \tfrac{1}{4}\alpha^2(1-\phi_3^2)\,, \tag{7.9}$$

where $\mathcal{L}_0$ is the Lagrangian for the unmodified O(3) model, given by (7.5), and α is a real constant. In terms of the complex field u,

$$\mathcal{L}_Q = \mathcal{L}_0 + \frac{\alpha^2 u\bar{u}}{(1+u\bar{u})^2}\,. \tag{7.10}$$

The corresponding equation of motion is

$$\partial_\mu\partial^\mu u - \frac{2\bar{u}\,\partial_\mu u\,\partial^\mu u}{1+u\bar{u}} - \frac{\alpha^2 u(1-u\bar{u})}{1+u\bar{u}} = 0\,. \tag{7.11}$$

The conserved charge Q arises from the invariance of $\mathcal{L}_Q$ under the multiplication of u by a global phase. Once again, the form of the potential has been carefully chosen to give simple analytic solutions, which may be constructed as follows using an analogue of the

Bogomolny equations. First of all, consider the following expressions for the total energy E, the topological charge k, and the Noether charge Q:

$$\begin{aligned} E &= \int \frac{\partial_i u\, \partial_i \bar{u}}{(1+u\bar{u})^2}\, d^2x + \int \frac{\partial_t u\, \partial_t \bar{u}}{(1+u\bar{u})^2}\, d^2x + \int \frac{\alpha^2 u\bar{u}}{(1+u\bar{u})^2}\, d^2x\,, \\ k &= \frac{i}{2\pi} \int \frac{\partial_x u\, \partial_y \bar{u} - \partial_y u\, \partial_x \bar{u}}{(1+u\bar{u})^2}\, d^2x\,, \\ Q &= i\alpha \int \frac{u\partial_t \bar{u} - \bar{u}\partial_t u}{(1+u\bar{u})^2}\, d^2x\,, \end{aligned} \tag{7.12}$$

where the subscript i labels the spatial coordinates x and y. Now consider the identity

$$\begin{aligned} \int (1+u\bar{u})^{-2} (\partial_i u \pm i\epsilon_{ij}\partial_j u)(\partial_i \bar{u} \mp i\epsilon_{ik}\partial_k \bar{u})\, d^2x \\ + \int (1+u\bar{u})^{-2} (\partial_t u \pm i\alpha u)(\partial_t \bar{u} \mp i\alpha \bar{u})\, d^2x \;\geq\; 0\,. \end{aligned} \tag{7.13}$$

Using (7.12), (7.13) may be recast in the simple form

$$E \geq 2\pi |k| + |Q|\,. \tag{7.14}$$

Given particular values of the (integer-valued) topological charge k, and the (real-valued) nontopological charge Q, (7.14) gives a lower bound on the total energy E, which is attained when

$$\partial_i u \pm i\epsilon_{ij}\partial_j u = 0 \qquad \text{and} \qquad \partial_t u \pm i\alpha u = 0\,. \tag{7.15}$$

The first of these conditions determines the spatial dependence of u and is familiar from the unmodified model; namely, if one writes $z = x + iy$ then u must be an (anti)-analytic function of z. The second part of (7.15) is new, and says simply that u should rotate in internal space with constant angular velocity α. Putting these two pieces together, one finds that the lower bound on the energy is attained if and only if

$$u(z,t) = u_0(z) \exp(\pm i\alpha t)\,, \tag{7.16}$$

where $u_0(z)$ an arbitrary (anti)-analytic function, *i.e.* a solution of the unmodified model. If $u_0(z)$ is chosen to be a k-instanton solution then $u(z,t)$ describes a system of k Q-lumps.

It only remains to check that the configurations (7.16) do indeed satisfy the field equation (7.11), but this is easily done. Moreover, it is also clear that the corresponding energy densities are independent of time, which shows that there are no forces between isolated Q-lumps (in contrast to the skyrmions discussed above).

The explicit form of these new Q-lumps is so simple that it is almost deceptive. It stems from the existence of the modified Bogomolny bound (7.14), and this in turn is a consequence of the careful choice of potential in (7.10). If one changes the potential, the Bogomolny bound is lost, along with the simple analytic solutions; so it seems that the model we have constructed is very special indeed.

It is now natural to ask about the stability and scattering of Q-lumps. First of all, note that if $u(z,t)$ has unit topological charge, *i.e.* if

$$u_0(z) = \frac{\lambda}{z-\mu}\,, \tag{7.17}$$

then Q (and hence also E) is infinite. In other words, if one restricts attention to finite-energy solutions then the minimum topological charge allowed is two. Using a refinement of the techniques developed in chapter IV, it has been verified that Q-lumps with $k \geq 2$ are stable to radially symmetric perturbations. However, it seems that under more general perturbations they may undergo fission, in a manner similar to conventional Q-balls when the continuous symmetry group is nonabelian.[66] For example, a Q-lump of topological charge two may split into two Q-lumps, each with unit charge.

Answers to this question, and also to the wider issue of Q-lump scattering, could come from a slow-motion approximation, in which the evolution is restricted to the manifold of configurations that saturate (7.14). Suppose this manifold is denoted by $M_k'(Q)$. Clearly $M_k'(Q)$ is a submanifold of M_k (recall from chapter VI that M_k is the manifold of charge-k static solutions in the unmodified model). The simplest case is $k=2$, corresponding to

$$u_0(z) = \frac{2\beta z + \gamma}{z^2 + \epsilon}\,, \tag{7.18}$$

where β is real, but γ and ϵ may be complex. The condition that Q is finite forces β to vanish, and then the phase of γ becomes irrelevant (see section 6.3). In addition, the fact that Q is conserved places a further constraint on γ and ϵ, with the result that $M_2'(Q)$ has dimension two. Any work along these lines would be all the more interesting in view

of the fact that, up till now, investigations of Q-ball scattering[72] have been confined to (1+1) dimensions.

As a final suggestion, it would be interesting to look at the whole family of modified σ-models that are given by

$$(\eta^{\mu\nu} + V_\alpha \epsilon^{\mu\nu\alpha})\, \partial_\mu (J^{-1} \partial_\nu J) = 0\,, \tag{7.19}$$

where J lives in SU(2) and V_α is a constant vector in spacetime. We have already seen that if $V_\alpha = (0,1,0)$ then the model is integrable, and that if $V_\alpha = (0,0,0)$ then one has the conventional SU(2) chiral model. It is natural to consider models corresponding to intermediate values of V. In particular, one may ask how the scattering properties change as V_α changes from $(0,0,0)$ to $(0,1,0)$, *i.e.* as the model changes from being Lorentz invariant to being integrable. The static solutions are always the same, regardless of the value of V, and so a slow-motion analysis may well be feasible, especially if the model contains no negative modes.

Perhaps such an investigation would give a valuable insight into the apparent nonexistence of (2+1)-dimensional soliton theories that are both Lorentz invariant *and* integrable. It might also establish a link between the geodesic approximation and the inverse scattering transform (IST), bearing in mind that, in general, the spectrum of eigenvalues in the IST contains both a continuous piece, corresponding to radiative components, and a discrete piece, corresponding to the solitonic components. One could think of the geodesic approximation as an inferior (and in some cases inadequate) substitute for the IST, in which one ignores all radiative effects and evolves only the solitonic components. This point of view is apparent in reference [73], which contains a comprehensive review of soliton dynamics in nearly integrable systems in (1+1) dimensions.

Appendix A

The main purpose of this appendix is to provide a motivation for the integrable chiral model of chapter III, by showing that it may be obtained as a reduction of the self-dual Yang-Mills (self-duality) equations in (2+2) dimensions (by (2+2) dimensions it is meant that the spacetime metric has signature $(++--)$). In addition, we take the opportunity to mention some other reductions of the self-duality equations, especially reductions from spacetime signature $(++++)$, *i.e.* from fields defined in (4+0) dimensions. For example, the solutions of the simplest reduction from (4+0) dimensions to (3+0) are the BPS monopoles, which were discussed in chapter VI. It seems that the self-duality equations and their generalizations occupy a central role in the theory of integrable systems,[74] in the sense that many (and maybe all) integrable equations can be obtained from them by some suitable reduction.

Briefly, the self-duality equations take the following form. A more complete treatment may be found in any of the standard texts, for example reference [23]. One has a gauge group G, with Lie algebra g, and a gauge potential A, which is a g-valued 1-form transforming as a connection under the action of the gauge group. The gauge field F (interpreted geometrically as a curvature) is a g-valued 2-form, defined by

$$F = dA + A \wedge A\,, \tag{A.1}$$

where d is the usual exterior derivative. In general, the covariant derivative of a matrix-valued p-form V is given by

$$DV = dV + A \wedge V - (-)^p V \wedge A\,. \tag{A.2}$$

The full Yang-Mills equations are simply

$$D^*F = 0\,, \tag{A.3}$$

where * is the Hodge duality operator. Being nonlinear, they are very difficult to solve,

but progress may be made by making use of the Bianchi identity, namely

$$\begin{aligned} DF &= dF + A \wedge F - F \wedge A \\ &= d(A \wedge A) + A \wedge dA - dA \wedge A \\ &= 0 . \end{aligned} \tag{A.4}$$

From (A.4), it is clear that a particular set of solutions to (A.3) consists of those gauge fields which satisfy

$$^*F = \lambda F , \tag{A.5}$$

where λ is a (possibly complex) constant. Equations (A.5) are the self-duality equations, and are much easier to solve than the full Yang-Mills equations.

It turns out that the possible values of λ depend on the spacetime signature. Applying the Hodge operator to (A.5) yields

$$^{**}F = \lambda {}^*F = \lambda^2 F , \tag{A.6}$$

and now one can use the standard results concerning the action of the repeated Hodge dual, namely that for a p-form V defined on an n-dimensional spacetime,

$$^{**}V = (-)^{p(n-p)+s} V , \tag{A.7}$$

where s is the number of minus signs in the spacetime signature. The conclusion is that when s is even, the allowed values of λ are ± 1, while for s odd they are $\pm i$.

The above discussion applies to arbitrary spacetime geometries and arbitrary gauge groups; but for the rest of the appendix, attention is restricted to flat four-dimensional spacetimes and to the gauge group SU(2), for which the individual components of the gauge potential must be real. This in turn means that λ must be real. In other words, one is restricted to signatures $(+{+}{+}{+})$, which is essentially the same as $(-{-}{-}{-})$, and

$(++--)$. The signature $(+++-)$ is precluded, although it *would* be allowed if the gauge group were changed to (say) $\mathrm{SL}(2,\mathbb{C})$. By convention, fields which satisfy

$$^*F = F \tag{A.8}$$

are called self-dual and those which satisfy

$$^*F = -F \tag{A.9}$$

are called anti-self-dual. In (4+0) dimensions the finite-action configurations of these types are known as instantons and anti-instantons respectively.

To begin with, we shall discuss the reduction from the self-dual Yang-Mills equations to the Bogomolny equation for monopoles. Consider a spacetime with signature $(++++)$ and coordinates (x^1, x^2, x^3, x^4); then the gauge potential may be written $A = A_i dx^i$, where i takes values 1, 2, 3 and 4. The simplest reduction is to take the fields to be independent of one coordinate, say by setting $\partial_4 = 0$. Now write

$$A = \tilde{A} + \Phi, \tag{A.10}$$

where

$$\tilde{A} = A_1 dx^1 + A_2 dx^2 + A_3 dx^3 \quad \text{and} \quad \Phi = A_4 dx^4. \tag{A.11}$$

Suppose that the full four-dimensional spacetime is denoted by S (here S is simply $\mathbb{R}^4$) and that $\tilde{S}$ is the three-dimensional subspace obtained by dropping the coordinate x^4. We shall use a tilde to denote any object defined on $\tilde{S}$ rather than S (note that $\tilde{d} = d$, since $\partial_4 = 0$). From the definition of the gauge field,

$$\begin{aligned} F &= d\tilde{A} + d\Phi + (\tilde{A} + \Phi) \wedge (\tilde{A} + \Phi) \\ &= \tilde{F} + \tilde{d}\Phi + \tilde{A} \wedge \Phi + \Phi \wedge \tilde{A} \\ &= \tilde{F} + \tilde{D}\Phi. \end{aligned} \tag{A.12}$$

In this way, F is decomposed into two pieces. The first, $\tilde{F}$, has no component involving dx^4; conversely, dx^4 appears in *every* component of $\tilde{D}\Phi$. Under the action of the Hodge dual, the two pieces are interchanged, so that $^*\tilde{F} = \tilde{D}\Phi$ and $^*\tilde{D}\Phi = \tilde{F}$. We can remove

all reference to x^4 by writing $\phi \equiv A_4$ and letting the Hodge dual act in $\tilde{S}$, so that it now takes 1-forms to 2-forms and vice-versa. It is conventional to define the 1-form $\tilde{B}$ to be the dual of $\tilde{F}$; then taking a little care with signs, it is easy to see that (A.8) becomes

$$\tilde{B} = \tilde{D}\phi . \tag{A.13}$$

This is the Bogomolny equation for monopoles. Note that the component of the gauge field corresponding to the redundant coordinate x^4 is now playing the role of the Higgs field.

A vast literature has been generated by the search for solutions to (A.8) (instantons) and (A.13) (monopoles). In both cases, the finite-energy configurations are classified by topological charge. Historically, the instantons were constructed before the monopoles. Indeed, for several years there were no examples of monopoles with charges greater than one, but eventually these solutions were found using techniques derived from the earlier work on instantons. A good review of monopoles, instantons, and the methods used to construct them may be found in reference [75], although it was written before the higher-charge monopoles were discovered.

To look at other reductions of the self-duality equations, it is convenient to use an alternative formalism, originally put forward by Yang,[76] in which one defines new spacetime coordinates, y, $\bar{y}$, z and $\bar{z}$, in such a way that the spacetime metric is

$$ds^2 = dy\, d\bar{y} + dz\, d\bar{z} . \tag{A.14}$$

The overbar *may* denote complex conjugation, but equally may not, depending on the spacetime signature.

$$\text{If the signature is } (++++) \text{ then } \begin{cases} y = x_1 + ix_2 , & z = x_3 + ix_4 , \\ \bar{y} = x_1 - ix_2 , & \bar{z} = x_3 - ix_4 . \end{cases} \tag{A.15}$$

$$\text{If the signature is } (++--) \text{ then } \begin{cases} y = x_1 + x_2 , & z = x_3 + x_4 , \\ \bar{y} = x_1 - x_2 , & \bar{z} = x_3 - x_4 . \end{cases} \tag{A.16}$$

One also introduces new components for the gauge potential, as follows.

$$\text{If the signature is } (++++) \text{ then } \begin{cases} A_y = A_1 + iA_2\,, & A_z = A_3 + iA_4\,, \\ A_{\bar{y}} = A_1 - iA_2\,, & A_{\bar{z}} = A_3 - iA_4\,. \end{cases} \tag{A.17}$$

$$\text{If the signature is } (++--) \text{ then } \begin{cases} A_y = A_1 + A_2\,, & A_z = A_3 + A_4\,, \\ A_{\bar{y}} = A_1 - A_2\,, & A_{\bar{z}} = A_3 - A_4\,. \end{cases} \tag{A.18}$$

The key to Yang's formalism is to write these new components in terms of two 2×2 matrices, D and $\bar{D}$, which may be taken to have unit determinant:

$$\begin{aligned} A_y &= D^{-1}\partial_y D\,, & A_z &= D^{-1}\partial_z D\,, \\ A_{\bar{y}} &= \bar{D}^{-1}\partial_{\bar{y}}\bar{D}\,, & A_{\bar{z}} &= \bar{D}^{-1}\partial_{\bar{z}}\bar{D}\,. \end{aligned} \tag{A.19}$$

The advantage of doing this is apparent when one considers that the self-duality equations are essentially three equations, equating pairs of the six independent components of F (recall that F is an antisymmetric object with two indices, each taking values 1, 2, 3 and 4). By writing A in the form (A.19), two of these three equations are automatically satisfied, and the third takes the form

$$\partial_{\bar{y}}(J^{-1}\partial_y J) + \partial_{\bar{z}}(J^{-1}\partial_z J) = 0\,, \tag{A.20}$$

where

$$J \equiv D\bar{D}^{-1}\,. \tag{A.21}$$

The similarity with the chiral field equations is now emerging; but before proceeding, we should mention a subtle difference between the signatures $(++++)$ and $(++--)$, which stems from the fact that the A_i $(i = 1, 2, 3, 4)$ may be thought of as 2×2 anti-hermitian matrices. In $(++++)$, this means that $A_y^\dagger = -A_{\bar{y}}$, so relating D and $\bar{D}$ via (A.19). Given a particular gauge potential, $\bar{D}$ is not uniquely determined in terms of D, but it is easy to see that one possible choice is always

$$\bar{D} = (D^\dagger)^{-1}\,. \tag{A.22}$$

This has the result that J is hermitian (in addition to having unit determinant).

It should be stressed that (A.22) applies only to signature (++++). In contrast, for signature (++−−), A_y, $A_{\bar{y}}$, A_z and $A_{\bar{z}}$ are themselves antihermitian; so D and $\tilde{D}$, and hence also J, may be taken to live in SU(2). Concentrating on this case, since it is of particular relevance to the σ-models, we now exhibit a few particular reductions, including the integrable chiral model of chapter III. First of all, set $\partial_2 = \partial_4 = 0$; then (A.20) becomes

$$\partial_i(J^{-1}\partial_i J) = 0\,, \tag{A.23}$$

where i takes values 1 and 2. This is simply the SU(2) chiral field equation in (2+0) dimensions. If J is restricted to lie on the equator of SU(2) then one has the O(3) σ-model.

The integrable chiral model comes from setting $\partial_4 = 0$ and relabelling the other three coordinates so that

$$\begin{aligned} x_1 &\to y\,, \\ x_2 &\to t\,, \\ x_3 &\to x\,. \end{aligned} \tag{A.24}$$

Equation (A.20) then reads

$$(\partial_y - \partial_t)\big(J^{-1}(\partial_y + \partial_t)J\big) + \partial_x(J^{-1}\partial_x J) = 0\,, \tag{A.25}$$

or alternatively

$$\begin{aligned} \partial_x(J^{-1}\partial_x J) + \partial_y(J^{-1}\partial_y J) - \partial_t(J^{-1}\partial_t J) \\ + \partial_y(J^{-1}\partial_t J) - \partial_t(J^{-1}\partial_y J) = 0\,, \end{aligned} \tag{A.26}$$

which is precisely the equation studied in chapter III.

Of course, there are many more possibilities. Mason and Sparling[40] have described how to obtain the Korteweg-de Vries and nonlinear Schrödinger equations; and yet more reductions are pointed out in reference [74]. All the equations obtained in this way are integrable, by virtue of the fact that the self-duality equations from which they are derived are themselves integrable. Indeed, their solutions can often be found by tailoring the techniques used to construct instantons.

Appendix B

This appendix fills in the mathematical details that were omitted from chapter IV. There are two main tasks: firstly we must show that if quantities λ_n are defined for lattice instantons according to (4.25) then they have a finite limit λ as n tends to infinity. Second, we have to derive equation (4.27), which relates λ to h_{n_0} for some large n_0. In fact, (4.27) will arise as a corollary of the proof that λ exists. Along the way, we shall need several intermediate results (lemmas 1–3), which allow us to pin down the asymptotic behaviour of the discrete Bogomolny relations.

First of all, we shall introduce another first-order difference relation, which is closely related to the Bogomolny relations (4.20) in the following sense. Take ϵ small and look for a solution to (4.20) of the form

$$h_n = \frac{1 - a_n\epsilon}{1 + a_n\epsilon} = 1 - 2a_n\epsilon + \mathcal{O}(\epsilon^2)\,. \tag{B.1}$$

To leading order in ϵ, (4.20) becomes

$$a_{n+1} = a_n\Big(1 + \frac{2N}{n}\Big) \qquad (n \geq 1)\,, \tag{B.2}$$

which has general solution

$$a_n = \frac{a_1}{(2N)!}\,\frac{(n + 2N - 1)!}{(n-1)!}\,. \tag{B.3}$$

This provides the motivation for defining a new set of lattice quantities,

$$h'_n \equiv \frac{1 - a_n}{1 + a_n}\,, \tag{B.4}$$

which, it is easily checked, are the general solution of the difference equation

$$h'_{n+1} = \frac{h'_n - (N/n)(1 - h'_n)}{1 + (N/n)(1 - h'_n)}\,. \tag{B.5}$$

This difference relation, with its explicit solution, is the key to the whole proof. Assume that a_1 is chosen to be positive. Then $a_n > 0$ for all $n \geq 1$; moreover $a_{n+1} > a_n$ and $a_n \to \infty$ as $n \to \infty$. So from (B.4) it is seen that $\{h'_n\}$ has the following properties in common with $\{h_n\}$:

1. $|h'_n| < 1$.
2. $h'_{n+1} < h'_n$.
3. $h'_n \to -1$ as $n \to \infty$.

It is now useful to define counterparts to the λ_n^2 for the new quantities h'_n:

$$\lambda_n'^2 \equiv \left(\frac{1+h'_n}{1-h'_n}\right) n^{2N} = \frac{n^{2N}}{a_n}, \tag{B.6}$$

where the last equality follows from (B.4). Using the explicit form of a_n given in (B.3), one finds that

$$\lambda_n'^2 \to \lambda'^2 = \frac{(2N)!}{a_1} \qquad \text{as } n \to \infty. \tag{B.7}$$

The required sequence of lemmas may now be constructed.

Lemma 1. Suppose that $\{h'_n\}$ and $\{h''_n\}$ both satisfy (B.5) and that $h'_n > h''_n$ for some n. Then $h'_{n+1} > h''_{n+1}$.

Proof. Let $D = \left(1 + \frac{N}{n}(1-h'_n)\right)\left(1 + \frac{N}{n}(1-h''_n)\right) > 0$; then (B.5) implies

$$\begin{aligned} h'_{n+1} - h''_{n+1} &= \frac{1}{D}\left(\left(h'_n - \frac{N}{n}(1-h'_n)\right)\left(1 + \frac{N}{n}(1-h''_n)\right) - (h'_n \leftrightarrow h''_n)\right) \\ &= \frac{1}{D}(h'_n - h''_n)\left(1 + \frac{2N}{n}\right) \\ &> 0. \quad \blacksquare \end{aligned}$$

Lemma 2. Suppose that $\{h_n\}$ satisfies the Bogomolny relations (4.20), that $\{h'_n\}$ satisfies (B.5) and also that $h'_n = h_n$ for some n. Then $h'_{n+1} > h_{n+1}$.

Proof. Let $D = 1 + \frac{N}{n}(1-h_n) > 0$; then (4.20) and (B.5) imply

$$\begin{aligned} h'_{n+1} - h_{n+1} &= \frac{1}{D}\left(h_n - \frac{N}{n}(1-h_n) - \left(1 + \frac{N}{n}(1-h_n)\right)\left(h_n - \frac{N}{n}(1-h_n^2)\right)\right) \\ &= \frac{1}{D}\frac{N^2}{n^2}(1-h_n)(1-h_n^2) \\ &> 0. \quad \blacksquare \end{aligned}$$

Lemma 3. $\{\lambda_n^2\}$ is bounded.

Proof. Lemmas 1 and 2 together show that if $h'_{n_0} = h_{n_0}$ for some n_0 then $h'_n > h_n$ for all $n > n_0$. Therefore $\lambda'^2_n > \lambda_n^2$ for $n > n_0$. But, given $\delta > 0$, $\lambda'^2_n < \lambda'^2 + \delta$ for sufficiently large n. Clearly $\{\lambda_n^2\}$ is bounded below by zero, and so

$$0 \leq \lambda_n^2 \leq \lambda'^2$$

for sufficiently large n. ∎

The final stages of the proof now follow. From (4.20)

$$1 - h_{n+1} = (1 - h_n)\big(1 + \tfrac{N}{n}(1 + h_n)\big)\,,$$
$$1 + h_{n+1} = (1 + h_n)\big(1 - \tfrac{N}{n}(1 - h_n)\big)\,.$$

Dividing one of these equations by the other, and using (4.25),

$$\frac{\lambda_{n+1}^2}{\lambda_n^2} = \Big(1 + \frac{1}{n}\Big)^{2N} \frac{1 - (N/n)(1 - h_n)}{1 + (N/n)(1 + h_n)}\,. \tag{B.8}$$

The fact that $\{\lambda_n^2\}$ is bounded, together with the form of (4.25), means that h_n must approach -1 at least as fast as $1/n^{2N}$, *i.e.* for sufficiently large n, $h_n = -1 + \mathcal{O}(n^{-2N})$. Therefore, expanding (B.8) in powers of $1/n$:

$$\frac{\lambda_{n+1}^2}{\lambda_n^2} = \left(1 + \frac{2N}{n} + \frac{N(2N-1)}{n^2}\right)\left(1 - \frac{2N}{n}\right) + \mathcal{O}(n^{-3})$$

$$= 1 - \frac{N(2N+1)}{n^2} + \mathcal{O}(n^{-3})\,, \tag{B.9}$$

which shows that $\{\lambda_n^2\}$ is monotonically decreasing for sufficiently large n. Since $\{\lambda_n^2\}$ is bounded below by zero, it must tend to a finite limit λ^2. ∎

Having shown that λ is well defined, we now carry on to show how equation (B.9) leads to an approximate value for λ in terms of h_{n_0}. By definition

$$\lambda^2 = \lambda_{n_0}^2 \prod_{n=n_0}^{\infty} \frac{\lambda_{n+1}^2}{\lambda_n^2}\,. \tag{B.10}$$

Take logarithms and choose n_0 sufficiently large so that for $n \geq n_0$ the $\mathcal{O}(n^{-3})$ terms in (B.9) are small compared with $\mathcal{O}(n^{-2})$. Then

$$\log \lambda^2 \approx \log \lambda_{n_0}^2 - \sum_{n=n_0}^{\infty} \frac{N(2N+1)}{n^2}$$

$$\approx \log \lambda_{n_0}^2 - N(2N+1)\left(\frac{\pi^2}{6} - \sum_{n=1}^{n_0-1} \frac{1}{n^2}\right), \tag{B.11}$$

where in the last step we have used the result

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}. \tag{B.12}$$

Introducing the notation

$$S_0 \equiv \frac{\pi^2}{6} - \sum_{n=1}^{n_0-1} \frac{1}{n^2}, \tag{B.13}$$

we get

$$\lambda^2 \approx \lambda_{n_0}^2 \exp\bigl(-S_0 N(2N+1)\bigr). \tag{B.14}$$

As it stands, (B.14) is of limited use because it still involves $\lambda_{n_0}^2$. It is much better to rewrite it by relating $\lambda_{n_0}^2$ to λ'^2. Using (B.6) and then (B.3) one finds

$$\frac{\lambda_{n+1}'^2}{\lambda_n'^2} = \left(1+\frac{1}{n}\right)^{2N} \frac{a_n}{a_{n+1}}$$

$$= \left(1+\frac{1}{n}\right)^{2N} \left(1+\frac{2N}{n}\right)^{-1}$$

$$= 1 + \frac{N(2N-1)}{n^2} + \mathcal{O}(n^{-3}). \tag{B.15}$$

If we again choose n_0 sufficiently large so that the $\mathcal{O}(n^{-3})$ terms are small then

$$\lambda'^2 \approx \lambda_{n_0}'^2 \exp\bigl(S_0 N(2N-1)\bigr). \tag{B.16}$$

But now recall that $h_{n_0} = h'_{n_0}$ and so $\lambda_{n_0} = \lambda'_{n_0}$. Hence comparison of (B.14) and (B.16)

yields

$$\lambda^2 \approx \lambda'^2 \exp(-4N^2 S_0)\,. \tag{B.17}$$

The final step involves rewriting λ'^2 in terms of h_{n_0}. Recall that in equation (B.7), λ'^2 was given in terms of a_1, but it may equally be expressed in terms of a_{n_0} (or h_{n_0}) by using (B.3), (B.4) and the fact that $h_{n_0} = h'_{n_0}$:

$$\lambda'^2 = \frac{(n_0 + 2N - 1)!}{a_{n_0}(n_0 - 1)!} = \frac{(n_0 + 2N - 1)!}{(n_0 - 1)!}\left(\frac{1 + h_{n_0}}{1 - h_{n_0}}\right). \tag{B.18}$$

So finally we obtain equation (4.27):

$$\lambda^2 \approx \frac{(n_0 + 2N - 1)!}{(n_0 - 1)!}\left(\frac{1 + h_{n_0}}{1 - h_{n_0}}\right) \exp(-4N^2 S_0)\,. \tag{B.19}$$

Appendix C

This appendix contains listings of the computer programs that were used to obtain some of the results in this thesis. The first, called CP1 (on pages 144 to 152), is an implementation of the topology-based evolution scheme that was developed in chapter IV and used in chapter V. The second, called CP1SLOW (on pages 153 to 159), applies the geodesic approximation discussed in chapter VI to the general problem of soliton scattering in the O(3) model, and makes extensive use of the NAG (Numerical Algorithms Group) routines.

Both programs are written in standard FORTRAN 77. They are designed to be reasonably efficient, but not at the expense of clarity. It is not intended to provide any documentation other than that contained in the listings themselves.

However, it is worth mentioning the behaviour of each numerical procedure as the soliton width approaches zero. In the program CP1, the time evolution takes place in a subroutine called STEP (page 150), which at one point takes the square root of a complicated argument. It seems that if the soliton width becomes smaller than some critical value then this argument becomes negative, causing the program to "crash". The critical width corresponds to a soliton occupying essentially only one lattice site, and is probably related to the minimum width of instanton that the lattice can support, as discussed in chapter IV.

Program CP1SLOW behaves somewhat differently. The values of γ and ϵ (see chapter VI) are printed at equally spaced time steps. If the width of either soliton becomes zero, then the program does not crash, but the next time step is never reached, no matter how much computer time one allocates. To be sure that a singularity has been found, one should run the job again, with an increased allocation. This is not entirely satisfactory (in addition to being expensive on computer time), but it is difficult to see how things might be improved, since the cause of this behaviour is presumably hidden deep inside the NAG routines. It certainly does not seem that the appearance of a singularity in the energy density corresponds to anything so simple as a vanishing denominator in the field equations (6.16).

```
      PROGRAM CP1
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
      IMPLICIT INTEGER*4 (I-N)

C     ************************************************************************
C     ****    R A D I A L L Y   S Y M M E T R I C   C P 1   M O D E L    ****
C     ************************************************************************

C     NMESH is the number of points in the radial mesh, and NCALC is the
C     number of those points used to generate the output (potential energy,
C     kinetic energy etc).

      PARAMETER ( NMESH = 6000, NCALC = 1000 )


C     Specify the topological sector and the time step for the evolution.

      PARAMETER ( NCH = 1 )
      PARAMETER ( TSTEP = 0.5D0 )


C     The basic output is printed after every NPRINT time steps up to a
C     maximum time TMAX.  More detailed output, handled by the subroutine
C     PLOT, is printed after every NPLOT time steps up to a maximum time
C     TPLOT.  Either type of output may be turned off by setting NPRINT
C     or NPLOT to -1.

      PARAMETER ( NPRINT = 10, TMAX = 1.0D4 )
      PARAMETER ( NPLOT = -1, TPLOT = 1.6D2 )


C     The type of boundary condition is specified by IBTYPE (see the
C     subroutine BOUND); the type of initial configuration is specified
C     by IHTYPE (see the subroutine INIT).

      PARAMETER ( IBTYPE = 0 )
      PARAMETER ( IHTYPE = 2 )


C     The fields are all stored in the matrix F.  The first index labels
C     the three components SQRT(1-H**2), PSI and H; the second index
C     specifies the position in the radial mesh; and the third index
C     labels the three most recent time slices.

      REAL*8 F  ( 1:3, 0:NMESH+1, 0:2 )


C     The following data are used in the construction of the initial
C     configuration (see the subroutine INIT).

      DATA        TL,    NO,    N1,        A,        B,      XL    /
     1          0.0D0,   100,   175,   1.0D-3,    1.0D3,   5.0D1   /

      DATA        TP                                                /
     1         5.0D-4                                             /


*****************************************
*****   Set up required common blocks   *****
*****************************************
```

```
      COMMON / MODEL / TS, PP, NM, NC, NL
      COMMON / BTYPE / IB

      TS = TSTEP
      PP = 4.0D0 * DATAN(1.0D0)
      NM = NMESH
      NC = NCH
      NL = NCALC
      IB = IBTYPE


C     ***************************************
C     ****    M A I N    P R O G R A M    ****
C     ***************************************

      NOW = 0
      TIME = 0.0D0
      WRITE (6,100)
      CALL INIT ( IHTYPE, F, A, B, XL, TL, N0, N1, TP )
      CALL STEP ( F, NOW )
      IF ( NPRINT .NE. -1 ) CALL OUTPUT ( F, TIME, EN, EO, NOW )
      IF ( NPLOT  .NE. -1 ) CALL PLOT ( F, NOW, TSTEP )
      NSTEP = INT ( TMAX / TSTEP ) + 1
      DO 200 I = 1, NSTEP
          TIME = TIME + TSTEP
          CALL STEP ( F, NOW )


          IF ( NPRINT .NE. -1 .AND. MOD ( I, NPRINT ) .EQ. 0 )
     1         CALL OUTPUT ( F, TIME, EN, EO, NOW )
          IF ( NPLOT .NE. -1 .AND. MOD ( I, NPLOT ) .EQ. 0 .AND.
     1         TIME .LT. TPLOT ) CALL PLOT ( F, NOW, TSTEP )


  200 CONTINUE
  100 FORMAT (//'Radially symmetric CP1 MODEL in (2+1) dimensions'/)
      WRITE (6,*)
      END


*************************************************************************
*****   Specification of initial configuration and perturbation   *****
*************************************************************************

      SUBROUTINE INIT ( IHTYPE, F, A, B, XL0, TL, N0, N1, TP )
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
      IMPLICIT INTEGER*4 (I-N)
      COMMON / MODEL / TSTEP, PI, NMESH, NCH
      COMMON / BTYPE / IBTYPE
      REAL*8 F ( 1:3, 0:NMESH+1, 0:2 )


C     The option IHTYPE=0 is used for one-off configurations.  It is set up
C     here to take initial data from the continuum instanton solutions, or
C     from the slow-motion approximation to a charge-two ring.

      IF ( IHTYPE .EQ. 0 ) THEN
          WRITE (6,10)
   10     FORMAT(1X,'One-off initial data specification'/)
```

```
            DO 20 J = 0, 1
                T = DFLOAT ( J-1 ) * TSTEP
                GAM = DSQRT(A) * DSQRT ( B**2 + T**2 )
                DO 30 I = 1, NMESH+1
                    R = DFLOAT ( I )
                    DEN = GAM**4 + R**4
C                   F(1,I,J) = 2.0D0 * XL0 * R / ( XL0**2 + R**2 )
C                   F(3,I,J) = ( XL0**2 - R**2 ) / ( XL0**2 + R**2 )
C                   F(2,I,J) = 0.0D0
                    F(1,I,J) = 2.0D0 * GAM**2 * R**2 / DEN
                    F(3,I,J) = ( GAM**4 -R**4 ) / DEN
                    F(2,I,J) = DATAN ( 2.0D0*B*T / ( B**2 - T**2 ) )
   30           CONTINUE
   20       CONTINUE


C       The option IHTYPE=1 sets up static instanton fields.  The parameters
C       used are NCH (the topological charge), N0 (see chapter IV) and XL0 (the
C       instanton width).

        ELSE IF ( IHTYPE .EQ. 1 ) THEN
            WRITE (6,100) NCH
  100       FORMAT(1X,'Single charge ',I1,' static instanton'/)
            PS = 0.0D0
            DO 101 J = 0, 1
                CALL STATIC ( J, N0, XL0, PS, F )
  101       CONTINUE


C       The option IHTYPE=2 applies perturbations to an instanton configuration.
C       In addition to NCH, N0 and XL0, there are parameters to specify the
C       perturbation.  TL and TP specify its size; N1 (together with N0)
C       specifies its shape.

        ELSE IF ( IHTYPE .EQ. 2 ) THEN
            WRITE (6,300) NCH
  300       FORMAT (1X,'Charge ',I1,' perturbed soliton'/)


C       The procedure comes in two parts.  First, the perturbation given
C       by TL and TP is constructed in timeslice J=1, taking no regard of
C       the perturbation shape.

            DO 302 J = 0, 1
                IF ( J .EQ. 0 ) XL = XL0
                IF ( J .EQ. 1 ) XL = XL0 + TSTEP * TL
                PS = TP * DFLOAT ( J ) * TSTEP
                CALL STATIC ( J, N0, XL, PS, F )
  302       CONTINUE


C       Now the shape of the perturbation is taken into account.  The fields at
C       timeslice J=1 are taken to be a linear combination of those calculated
C       in the previous section.  First, there is no perturbation beyond N=N1.

            DO 305 I = N1+1, NMESH+1
            F(3,I,1) = F(3,I,0)
            F(2,I,1) = F(2,I,0)
            F(1,I,1) = F(1,I,0)
  305       CONTINUE
```

```
C       In the central region, the shape is that of a linear ramp.

            DO 307 I = N0, N1
                 P1 = DFLOAT ( N1-I )  / DFLOAT ( N1-N0 )
                 P2 = 1.0D0 - P1
                 F(3,I,1) = P1 * F(3,I,1) + P2 * F(3,I,0)
                 F(2,I,1) = P1 * F(2,I,1) + P2 * F(2,I,0)
                 F(1,I,1) = DSQRT ( 1.0D0 - F(3,I,1)**2 )
  307       CONTINUE


C       The option IHTYPE=3 is used to set up properly discretized initial data
C       for a slowly moving charge-two ring.  The parameters used are A, B (see
C       chapter IV), and also N0.  Ensure that NCH is set to 2 !!!

        ELSE IF ( IHTYPE .EQ. 3 ) THEN
            WRITE (6,200)
  200       FORMAT(1X,'Slowly moving charge 2 ring'/)
            DO 202 J = 0, 1
                T  = DFLOAT ( J-1 ) * TSTEP
                XL = A * ( B**2 + T**2 )
                PH = DATAN ( 2.0D0*B*T / ( B**2 - T**2 ) )
                CALL STATIC ( J, N0, XL, PH, F )
  202       CONTINUE
        ENDIF


C       For any type of initial configuration, the fields at the origin must be
C       of the form (0,psi,1).  The following lines are included to be safe.

        DO 999 J = 0,2
            F(1,0,J) = 0.0D0
            F(2,0,J) = F(2,1,J)
            F(3,0,J) = 1.0D0
  999   CONTINUE


C       The last thing to do is to set the boundary values correctly.

        IF ( IBTYPE .NE. 0 ) CALL BOUND ( F, 0, IBTYPE )
        CALL BOUND ( F, 1, IBTYPE )
        CALL BOUND ( F, 2, IBTYPE )
        RETURN
        END


**********************************************
*****   Calculation of instanton fields   *****
**********************************************

        SUBROUTINE STATIC ( J, N0, XL, PS, F )
        IMPLICIT DOUBLE PRECISION (A-H,O-Z)
        IMPLICIT INTEGER*4 (I-N)
        COMMON / MODEL / TSTEP, PI, NMESH, NCH
        REAL*8 F ( 1:3, 0:NMESH+1, 0:2 )

        XCH = DFLOAT ( NCH )
        S0  = PI**2 / 6.0D0
```

```
      RO  = 1.0D0

C     This routine is called by several of the options in subroutine INIT.
C     It calculates the lattice instanton fields given by NO, XL and PS,
C     and then stores them in timeslice J.

      DO 100 I = 0, 2*NCH-1
      RO = RO * DFLOAT ( NO+I )
  100 CONTINUE
      DO 101 I = 1, NO-1
      SO = SO - 1.0D0 / DFLOAT ( I**2 )
  101 CONTINUE
      EP = XL**2 * DEXP ( 4.0D0 * XCH**2 * SO ) / RO


C     Calculate the fields at position NO first, and then the others.

      F(3,NO,J) = - ( 1.0D0 - EP ) / ( 1.0D0 + EP )
      F(2,NO,J) = PS
      F(1,NO,J) = DSQRT ( 1.0D0 - F(3,NO,J)**2 )
      DO 102 I = NO, 2, -1
          R = DFLOAT ( I-1 )
          Q = 1.0D0 + 4.0D0 * XCH * ( XCH/R + F(3,I,J) ) / R
          F(3,I-1,J) = R / 2.0D0 / XCH * ( DSQRT ( Q ) - 1.0D0 )
          F(2,I-1,J) = PS
          F(1,I-1,J) = DSQRT ( 1.0D0 - F(3,I-1,J)**2 )
  102 CONTINUE
      DO 103 I = NO, NMESH
          F(3,I+1,J) = F(3,I,J) - XCH * F(1,I,J)**2 / DFLOAT ( I )
          F(2,I+1,J) = PS
          F(1,I+1,J) = DSQRT ( 1.0D0 - F(3,I+1,J)**2 )
  103 CONTINUE
      RETURN
      END


**************************************************
*****   Summary output: energies and width   *****
**************************************************

      SUBROUTINE OUTPUT ( F, TIME, EN, EO, NOW )
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
      IMPLICIT INTEGER*4 (I-N)
      REAL*8 F ( 1:3, 0:NMESH+1, 0:2 ), KE, KEH, KEP
      COMMON / MODEL / TSTEP, PI, NMESH, NCH, NCALC


C     Check for incorrectly normalized fields - there should never be any !!

      DO 100 I = 0, NMESH+1
          F2 = F(1,I,NOW)**2 + F(3,I,NOW)**2
          IF ( DABS(F2-1.0D0) .GT. 1.0D-8 ) WRITE(6,*) I,F2
  100 CONTINUE


C     Calculate the total kinetic energy (KE), potential energy (PE) and the
C     width (WI).  KE is split into KEH, due to the time derivative of H,
C     and KEP, due to the time derivative of psi.
```

```
      KE  = 0.0D0
      PE  = 0.0D0
      WI  = 0.0D0
      KEH = 0.0D0
      KEP = 0.0D0
      NXT = MOD ( NOW+1, 3 )
      DO 200 I = 0, NCALC
          IF ( I .EQ. 0 ) AREA = 4.0D0
          IF ( I .NE. 0 ) AREA = 2.0D0 * DFLOAT ( I )
          PTE = EPT ( F, I, NOW, PTKH, PTKP, PTP )
          PTK = PTKH + PTKP
          KEH = KEH  + PTKH * AREA
          KEP = KEP  + PTKP * AREA
          KE  = KE   + PTK  * AREA
          PE  = PE   + PTP  * AREA
          WI  = WI   + PTE  * AREA * DFLOAT ( I )
  200 CONTINUE
      KEH = PI * KEH
      KEP = PI * KEP
      KE  = PI * KE
      PE  = PI * PE
      WI  = PI * WI
      EN  = KE + PE
      WI  = WI / EN


      IF ( TIME .EQ. 0.0D0 ) THEN
          E0 = EN
          WRITE (6,112) EN
          WRITE (6,113)
      ENDIF


C     Output statements.  It is useful to print F(3,NCALC+1,NOW) so that one
C     can calculate the discrete Bogomolny bound.

      WRITE  (6,111) TIME, KE, PE, WI, EN/E0
      WRITE  (6,114) KEP, F(3,NCALC+1,NOW)


  111 FORMAT (1X,F7.2,2X,3(D15.8,2X),D16.9)
  114 FORMAT (10X,D15.8,2X,D15.8)
  112 FORMAT (1X,'Initial energy = ',D15.8/)
  113 FORMAT (4X,'Time',
     1        11X,'Kin En',11X,'Pot En',12X,'Width',12X,'Rel En'/
     1        19X,'KE Psi',11X,'Bbound'/)
      RETURN
      END


*************************************************************
*****   Calculation of the discrete energy density    *****
*************************************************************

      DOUBLE PRECISION FUNCTION EPT ( F, I, NOW, KEH, KEP, PE )
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
      IMPLICIT INTEGER*4 (I-N)
      COMMON / MODEL / TSTEP, PI, NMESH, NCH
      REAL*8 F ( 1:3, 0:NMESH+1, 0:2 ), KEH, KEP
```

```
      NXT = MOD ( NOW+1, 3 )
      FH = 0.0D0
      FP = 0.0D0
      AT = TSTEP**2

      IF ( I .EQ. 0 ) THEN
          FR = DFLOAT ( NCH ) * ( 1.0D0 - F(3,1,NOW) )
      ELSE
          FR = DFLOAT ( NCH**2 ) * F(1,I,NOW)**2 / DFLOAT ( I**2 )
          FR = FR + ( F(3,I+1,NOW)-F(3,I,NOW) ) ** 2 / F(1,I,NOW)**2
          FR = FR + ( F(2,I+1,NOW)-F(2,I,NOW) ) ** 2 * F(1,I,NOW)**2
          FH = ( F(3,I,NXT)-F(3,I,NOW) ) ** 2 / F(1,I,NOW)**2 / AT
          FP = ( F(2,I,NXT)-F(2,I,NOW) ) ** 2 * F(1,I,NOW)**2 / AT
      ENDIF

      KEH = FH / 4.0D0
      KEP = FP / 4.0D0
      PE  = FR / 4.0D0
      EPT = KEH + KEP + PE
      RETURN
      END


*****************************************
*****   Single step time evolution   *****
*****************************************

      SUBROUTINE STEP ( F, NOW )
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
      IMPLICIT INTEGER*4 (I-N)
      REAL*8 F ( 1:3, 0:NMESH+1, 0:2 )
      COMMON / MODEL / TSTEP, PI, NMESH, NCH
      COMMON / BTYPE / IBTYPE

      NOW = MOD ( NOW+1, 3 )
      NXT = MOD ( NOW+1, 3 )
      LST = MOD ( NXT+1, 3 )
      XCH = DFLOAT ( NCH )
      AT  = TSTEP**2


C     Set up some useful shorthand notation.

      DO 100 I = 1, NMESH
          FI = DFLOAT ( I-1 ) / DFLOAT ( I )
          H  = F(3,I,NOW)
          HL = F(3,I-1,NOW)
          HR = F(3,I+1,NOW)
          HD = F(3,I,LST)
          P  = F(2,I,NOW)
          PL = F(2,I-1,NOW)
          PR = F(2,I+1,NOW)
          PD = F(2,I,LST)


C     First, do the evolution for psi, or F(2,I,J).

          A1 = ( HD**2 - 1.0D0 ) * ( P - PD )
          A2 = FI * AT * ( HL**2 - 1.0D0 ) * ( P - PL )
          PU = P - AT * ( P - PR ) + ( A1 - A2 ) / ( H**2 - 1.0D0 )
```

```
            F(2,I,NXT) = PU

C       Second, do the (more complicated) evolution for H, or F(3,I,J).

            B = ( H - HR ) * ( H*HR - 1.0D0 ) / F(1,I,NOW)**4
            IF ( I .EQ. 1 ) THEN
            B = B + XCH * ( XCH*H + 1.0D0 )
            ELSE
            B = B + XCH**2 * H / DFLOAT ( I**2 )
            B = B - FI * ( H - HL ) / ( 1.0D0 - HL**2 )
            ENDIF
            B = B * AT
            B = B + ( H - HD ) / ( 1.0D0 - HD**2 )
            B = B + H * ( AT * ( P - PR ) ** 2 - ( P - PU ) ** 2 )
            IF ( H .EQ. 0.0D0 ) THEN
            F(3,I,NXT) = B
            ELSE
            ROOT = ( 1.0D0 - H**2 ) * DSQRT ( 1.0D0 - 4.0D0 * H * B )
            F(3,I,NXT) = ( 1.0D0 + H**2 - ROOT ) / ( 2.0D0 * H )
            ENDIF
            F(1,I,NXT) = DSQRT ( 1.0D0 - F(3,I,NXT)**2 )
  100   CONTINUE


C       Finally, set the boundary values according to the chosen boundary cond.

        CALL BOUND ( F, NXT, IBTYPE )
        RETURN
        END


*****************************************************************
*****   Detailed output of the fields and energy density    *****
*****************************************************************

        SUBROUTINE PLOT ( F, NOW, TSTEP )
        IMPLICIT DOUBLE PRECISION (A-H,O-Z)
        IMPLICIT INTEGER*4 (I-N)
        REAL*8 F ( 1:3, 0:NMESH+1, 0:2 )
        COMMON / MODEL / T, PI, NMESH, NCH, NCALC


C       This routine is controlled by NPLOT, and can be used to output either
C       various components of the field, or one of the energy densities, etc.

        WRITE (9,*) NCALC
        DO 100 I = 0, NCALC
        XE = EPT ( F, I, NOW, XKH, XKP, XP )
        XK = XKH + XKP
        WRITE (9,*) XK, XP, XE
C       WRITE (9,*) F(2,I,NOW)
  100   CONTINUE
        RETURN
        END


*****************************************************
*****   Specification of the boundary condition   *****
*****************************************************
```

```
      SUBROUTINE BOUND ( F, NXT, IBTYPE )
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
      IMPLICIT INTEGER*4 (I-N)
      COMMON / MODEL / TSTEP, PI, NMESH, NCH
      REAL*8 F ( 1:3, 0:NMESH+1, 0:2 )


C     IBTYPE = 0  <->  Fixed boundary values.
C     IBTYPE = 1  <->  Zero gradient on the boundary.
C     IBTYPE = 2  <->  Charge NCH falloff on the boundary.

      IF ( IBTYPE .EQ. 2 ) THEN
      X = F(3,NMESH,NXT)
      Y = DFLOAT ( NCH ) / DFLOAT ( NMESH )
      F(3,NMESH+1,NXT) = X - ( 1.0D00 - X**2 ) * Y
      F(1,NMESH+1,NXT) = DSQRT ( 1.0D0 - F(3,NMESH+1,NXT)**2 )
      F(2,NMESH+1,NXT) = F(2,NMESH,NXT)
      ENDIF

      IF ( IBTYPE .EQ. 1 ) THEN
      F(3,NMESH+1,NXT) = F(3,NMESH,NXT)
      F(1,NMESH+1,NXT) = F(1,NMESH,NXT)
      F(2,NMESH+1,NXT) = F(2,NMESH,NXT)
      ENDIF

      IF ( IBTYPE .EQ. 0 ) THEN
      NOW = MOD ( NXT+2, 3 )
      F(3,NMESH+1,NXT) = F(3,NMESH+1,NOW)
      F(2,NMESH+1,NXT) = F(2,NMESH+1,NOW)
      F(1,NMESH+1,NXT) = F(1,NMESH+1,NOW)
      ENDIF


C     In addition, set the gradient of psi at the origin to zero.

      F(2,0,NXT) = F(2,1,NXT)
      RETURN
      END
```

```
      PROGRAM SLOW
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
      IMPLICIT INTEGER*4 (I-N)

C     ************************************************************************
C     ****    C P 1   G E O D E S I C   A P P R O X I M A T I O N    ****
C     ************************************************************************

C     N is the number of first-order equations to be solved.  There are four
C     coupled second-order equations, which are equivalent to eight first-
C     order equations.

      PARAMETER ( N = 8 )


C     Specify the value of beta.

      PARAMETER ( B = 0.0D0 )


C     Specify the initial values of C, D, E, F, and their time derivatives.

      PARAMETER ( C  =   1.0D1, D  = 0.0D0 )
      PARAMETER ( E  = -25.0D0, F  = 0.0D0 )
      PARAMETER ( CT =  -0.2D0, DT = 0.2D0 )
      PARAMETER ( ET =   1.0D0, FT = 0.0D0 )

      REAL*8 Y(N), W(23+21*N)
      EXTERNAL D02CBF, FCN, OUTPUT
      COMMON / MODEL / TSTEP, T1
      COMMON / PARAMS / BETA
      BETA = B * 2.0D0


C     Put the initial values into the correct places in the array Y(N).

      Y(1) = C
      Y(2) = D
      Y(3) = E
      Y(4) = F
      Y(5) = CT
      Y(6) = DT
      Y(7) = ET
      Y(8) = FT


C     T0 is the initial time (it may be different from zero if one wants to
C     restart a previous evolution).  T1 is the time at which the evolution
C     will stop, and TSTEP is the time interval at which the values of C, D,
C     E and F will be printed out.

      T0 = 0.0D0
      T1 = 1.5D2
      TSTEP = 0.5D0


C     The equations are evolved using the NAG routine D02CBF.

      TOL = 1.0D-6
      IFAIL = 0
```

```
      IRELAB = 0
      M = N
      WRITE (6,100) B, Y(1), Y(2), Y(3), Y(4)
      WRITE (6,101) Y(5), Y(6), Y(7), Y(8)
      WRITE (6,102) TOL
      WRITE (6,103)
      CALL D02CBF ( T0, T1, M, Y, TOL, IRELAB, FCN, OUTPUT, W, IFAIL )
      WRITE (6,*)
      CALL ERROR ( IFAIL, 3 )
      IF ( TOL .LT. 0.0D0 ) WRITE (6,104)


  100 FORMAT(/1X,'SLOW MOTION CP1 EVOLUTION'
     1       /1X,'------------------------'
     1      //1X,'Initial parameter values specified as follows :-'
     1       /1X,'Beta        = ',D16.9
     1       /1X,'Gamma       = ',D16.9,' + i ',D16.9
     1       /1X,'Epsilon     = ',D16.9,' + i ',D16.9)
  101 FORMAT( 1X,'Gamma dot   = ',D16.9,' + i ',D16.9
     1       /1X,'Epsilon dot = ',D16.9,' + i ',D16.9)
  102 FORMAT( 1X,'Error tol   = ',D16.9/)
  103 FORMAT(/4X,'Time',6X,'C=Re(gamma)',6X,'D=Im(gamma)',
     1        4X,'E=Re(epsilon)',4X,'F=Im(epsilon)'/)
  104 FORMAT(/1X,'Range of integration too short for TOL')
      END


*******************************
*****  Output routine   *****
*******************************

      SUBROUTINE OUTPUT ( TSOL, Y )
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
      IMPLICIT INTEGER*4 (I-N)
      PARAMETER (N=8)
      REAL*8 Y(N)
      COMMON / MODEL / TSTEP, T1
      WRITE (6,100) TSOL, ( Y(I), I = 1, 4 )
      WRITE (9) TSOL, ( Y(I), I = 1, 8 )
  100 FORMAT(1X,F7.2,3X,4(D14.7,3X))
      TSOL = TSOL + TSTEP
      RETURN
      END


*******************************
*****  Error handling   *****
*******************************

      SUBROUTINE ERROR ( IFAIL, IROUTE )
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
      IMPLICIT INTEGER*4 (I-N)

      IF ( IFAIL  .EQ. 0 ) GOTO 10
      IF ( IROUTE .EQ. 1 ) WRITE (6,100) IFAIL
      IF ( IROUTE .EQ. 2 ) WRITE (6,101) IFAIL
      IF ( IROUTE .EQ. 3 ) WRITE (6,102) IFAIL
      STOP
  100 FORMAT (/1X,'NAG error in routine D01AMF  -  IFAIL = ',I1)
  101 FORMAT (/1X,'NAG error in routine D01AJF  -  IFAIL = ',I1)
```

```
  102   FORMAT (/1X,'NAG error in routine D02CBF  -  IFAIL = ',I1)
   10   CONTINUE
        RETURN
        END


**********************************************
****    F U N C T I O N   L I B R A R Y    ****
**********************************************


***********************************
****    RINTEG ( NFUNC, Y )    ****
***********************************

C       On entry, NFUNC specifies one of 16 possible functions to be integrated.
C       Y(N) contains the current values of C, D, E, F etc, which may appear
C       as coefficients in the function NFUNC.  On exit, RINTEG contains the
C       integral of the function NFUNC over all space, evaluated using the
C       current values of Y(N).

        DOUBLE PRECISION FUNCTION RINTEG ( NFUNC, Y )
        IMPLICIT DOUBLE PRECISION (A-H,O-Z)
        IMPLICIT INTEGER*4 (I-N)
        PARAMETER ( LW = 2000, LIW = 252, N = 8 )
        REAL*8 W(LW), Y(N), YCOPY(N)
        INTEGER*4 IW(LIW)
        EXTERNAL F1
        COMMON / PARAMS / B, YCOPY, DUMMY, NFCOPY


C       Take a copy of the current values of Y(N) to be stored in COMMON.

        DO 10 I = 1, N
        YCOPY(I) = Y(I)
   10   CONTINUE
        NFCOPY = NFUNC


C       The radial integration is performed by the NAG routine D01AMF.

        EPSABS = 1.0D-8
        EPSREL = 1.0D-5
        INF = 1
        BOUND = 0.0D0
        IFAIL = 0
        CALL D01AMF ( F1, BOUND, INF, EPSABS,
     1  EPSREL, RESULT, ABSERR, W, LW, IW, LIW, IFAIL )
        CALL ERROR ( IFAIL, 1 )
        RINTEG = RESULT
        RETURN
        END


*****************************
****    F1 ( RCOPY )    ****
*****************************

C       This is the angular integration routine used by RINTEG.  On entry,
C       RCOPY contains a value of R, which is integrated over by RINTEG.
```

```
C       The angular integration is performed by the NAG routine D01AJF.

        DOUBLE PRECISION FUNCTION F1 ( RCOPY )
        IMPLICIT DOUBLE PRECISION (A-H,O-Z)
        IMPLICIT INTEGER*4 (I-N)
        PARAMETER ( LW = 2000, LIW = 252, N = 8 )
        REAL*8 W(LW), Y(N)
        INTEGER*4 IW(LIW)
        EXTERNAL F2
        COMMON / PARAMS / BETA, Y, R, NFUNC


C       Store the value of R in COMMON, so that F2(T) can access it.

        R = RCOPY


C       Integrate with respect to T over the range 0 to 2*PI.

        EPSABS = 1.0D-8
        EPSREL = 1.0D-5
        A = 0.0D0
        B = 2.0D0 * X01AAF(B)
        IFAIL = 0
        CALL D01AJF ( F2, A, B, EPSABS,
     1  EPSREL, RESULT, ABSERR, W, LW, IW, LIW, IFAIL )
        CALL ERROR ( IFAIL, 2 )
        F1 = RESULT
        RETURN
        END


************************
****    F2 ( T )    ****
************************

C       This routine calculates the integrand used by F1 ( RCOPY ).  It takes
C       the values of NFUNC and Y specified in RINTEG, along with the value of
C       RCOPY specified by F1, and computes the corresponding function value.

        DOUBLE PRECISION FUNCTION F2 ( T )
        IMPLICIT DOUBLE PRECISION (A-H,O-Z)
        IMPLICIT INTEGER*4 (I-N)
        PARAMETER (N=8)
        REAL*8 Y(N)
        COMMON / PARAMS / B, Y, R, NFUNC


C       All the functions are built out of the following five quantities.

        XC  = Y(1)  + B * R * DCOS(T)
        XD  = Y(2)  + B * R * DSIN(T)
        XE  = Y(3)  + R **2 * DCOS(2.0D0*T)
        XF  = Y(4)  + R **2 * DSIN(2.0D0*T)
        DEN = XC**2 + XD**2 + XE**2 + XF**2


C       Choose the function specified by NFUNC.

        IF ( NFUNC .EQ. 1 ) THEN
```

```
      F2 = (  XE**2 + XF**2 ) / DEN**2

      ELSE IF ( NFUNC .EQ. 2 ) THEN
      F2 = (  XC**2 + XD**2 ) / DEN**2

      ELSE IF ( NFUNC .EQ. 3 ) THEN
      F2 = ( -XC*XE - XD*XF ) / DEN**2

      ELSE IF ( NFUNC .EQ. 4 ) THEN
      F2 = (  XD*XE - XC*XF ) / DEN**2

      ELSE IF ( NFUNC .EQ. 5 ) THEN
      F2 = -4.0D0 * XC * ( XE**2+XF**2 ) / DEN**3

      ELSE IF ( NFUNC .EQ. 6 ) THEN
      F2 = -4.0D0 * XD * ( XE**2+XF**2 ) / DEN**3

      ELSE IF ( NFUNC .EQ. 7 ) THEN
      F2 = -4.0D0 * XE * ( XE**2+XF**2 ) / DEN**3 + 2.0D0*XE/DEN**2

      ELSE IF ( NFUNC .EQ. 8 ) THEN
      F2 = -4.0D0 * XF * ( XE**2+XF**2 ) / DEN**3 + 2.0D0*XF/DEN**2

      ELSE IF ( NFUNC .EQ. 9 ) THEN
      F2 = -4.0D0 * XC * ( XC**2+XD**2 ) / DEN**3 + 2.0D0*XC/DEN**2

      ELSE IF ( NFUNC .EQ. 10 ) THEN
      F2 = -4.0D0 * XD * ( XC**2+XD**2 ) / DEN**3 + 2.0D0*XD/DEN**2

      ELSE IF ( NFUNC .EQ. 11 ) THEN
      F2 = -4.0D0 * XE * ( XC**2+XD**2 ) / DEN**3

      ELSE IF ( NFUNC .EQ. 12 ) THEN
      F2 = -4.0D0 * XF * ( XC**2+XD**2 ) / DEN**3

      ELSE IF ( NFUNC .EQ. 13 ) THEN
      F2 = 4.0D0 * XC * ( XC*XE + XD*XF ) / DEN**3 - XE/DEN**2

      ELSE IF ( NFUNC .EQ. 14 ) THEN
      F2 = 4.0D0 * XD * ( XC*XE + XD*XF ) / DEN**3 - XF/DEN**2

      ELSE IF ( NFUNC .EQ. 15 ) THEN
      F2 = 4.0D0 * XE * ( XC*XE + XD*XF ) / DEN**3 - XC/DEN**2

      ELSE IF ( NFUNC .EQ. 16 ) THEN
      F2 = 4.0D0 * XF * ( XC*XE + XD*XF ) / DEN**3 - XD/DEN**2

      ENDIF
      F2 = R * F2
      RETURN
      END


*****************************************************
****    Computation of the evolution equations    ****
*****************************************************

      SUBROUTINE FCN ( T, Y, F )
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
      IMPLICIT INTEGER*4 (I-N)
```

```
      PARAMETER (N=8)
      REAL*8 Y(N), F(N)

      P  = RINTEG (  1, Y )
      Q  = RINTEG (  2, Y )
      R  = RINTEG (  3, Y )
      S  = RINTEG (  4, Y )
      PC = RINTEG (  5, Y )
      PD = RINTEG (  6, Y )
      PE = RINTEG (  7, Y )
      PF = RINTEG (  8, Y )
      QC = RINTEG (  9, Y )
      QD = RINTEG ( 10, Y )
      QE = RINTEG ( 11, Y )
      QF = RINTEG ( 12, Y )
      RC = RINTEG ( 13, Y )
      RD = RINTEG ( 14, Y )
      RE = RINTEG ( 15, Y )
      RF = RINTEG ( 16, Y )


C     Use the four linear relations between the various derivatives to
C     calculate the derivatives of S with respect to C, D, E and F.

      SC = PF - RD
      SD = RC - PE
      SE = RF - QD
      SF = QC - RE


C     Check for a zero denominator in the field equations.

      DEN = R**2 + S**2 - P*Q
      IF ( DEN .EQ. 0.0D0 ) THEN
      WRITE (6,100) P, Q, R, S
      STOP
      ENDIF
  100 FORMAT(/1X,'***  Zero denominator in evolution equations  ***'/
     1        6X,'P = ',D14.7/
     1        6X,'Q = ',D14.7/
     1        6X,'R = ',D14.7/
     1        6X,'S = ',D14.7/)


C     Now construct the right hand sides of the evolution equations.

      CTCT = Y(5)**2
      DTDT = Y(6)**2
      ETET = Y(7)**2
      FTFT = Y(8)**2
      CTDT = Y(5)*Y(6)
      CTET = Y(5)*Y(7)
      CTFT = Y(5)*Y(8)
      DTET = Y(6)*Y(7)
      DTFT = Y(6)*Y(8)
      ETFT = Y(7)*Y(8)


      AC = PC*CTCT - PC*DTDT
     1       + 2.0D0*PD*CTDT + 2.0D0*(RD+SC)*DTET + (2.0D0*RE-QC)*ETET
```

```
1         + 2.0D0*PE*CTET - 2.0D0*(RC-SD)*DTFT + (2.0D0*SF-QC)*FTFT
1         + 2.0D0*PF*CTFT + 2.0D0*(RF+SE)*ETFT


      AD = PD*DTDT - PD*CTCT
1         + 2.0D0*PC*CTDT - 2.0D0*(RD+SC)*CTET - (2.0D0*SE+QD)*ETET
1         + 2.0D0*PE*DTET + 2.0D0*(RC-SD)*CTFT + (2.0D0*RF-QD)*FTFT
1         + 2.0D0*PF*DTFT + 2.0D0*(RE-SF)*ETFT


      AE = QE*ETET - QE*FTFT
1         + 2.0D0*QC*CTET + 2.0D0*(RD-SC)*CTDT + (2.0D0*RC-PE)*CTCT
1         + 2.0D0*QD*DTET + 2.0D0*(RF-SE)*CTFT - (2.0D0*SD+PE)*DTDT
1         + 2.0D0*QF*ETFT - 2.0D0*(RE+SF)*DTFT


      AF = QF*FTFT - QF*ETET
1         + 2.0D0*QC*CTFT + 2.0D0*(RC+SD)*CTDT + (2.0D0*SC-PF)*CTCT
1         + 2.0D0*QD*DTFT - 2.0D0*(RF-SE)*CTET + (2.0D0*RD-PF)*DTDT
1         + 2.0D0*QE*ETFT + 2.0D0*(RE+SF)*DTET


      AC = 0.5D0 * AC
      AD = 0.5D0 * AD
      AE = 0.5D0 * AE
      AF = 0.5D0 * AF

      F(1) = Y(5)
      F(2) = Y(6)
      F(3) = Y(7)
      F(4) = Y(8)
      F(5) = ( Q*AC - S*AF - R*AE ) / DEN
      F(6) = ( Q*AD + S*AE - R*AF ) / DEN
      F(7) = ( S*AD + P*AE - R*AC ) / DEN
      F(8) = ( P*AF - S*AC - R*AD ) / DEN

      RETURN
      END
```

References

[1] R. A. Leese, *J. Math. Phys.* **30** (1989) 2072.

[2] R. A. Leese, *Phys. Rev.* **D40** (1989) 2004.

[3] R. A. Leese, M. Peyrard and W. J. Zakrzewski, Durham preprint DTP-89/27 (1989), to appear in *Nonlinearity*.

[4] R. A. Leese, Durham preprint DTP-90/01 (1990), submitted to *Nuclear Physics B*.

[5] M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, 1981).

[6] A. Nakamura, *Prog. Theo. Phys.* **Supplement 94** (1988) 195.

[7] P. Goddard and P. Mansfield, *Rep. Prog. Phys.* **49** (1986) 725.

[8] T. H. R. Skyrme, *Proc. Roy. Soc.* **A260** (1961) 127, *Nucl. Phys.* **31** (1962) 556.

[9] G. Holzwarth and B. Schwesinger, *Rep. Prog. Phys.* **49** (1986) 825.

[10] I. Dzyaloshinskii, A. Polyakov and P. Wiegmann, *Phys. Lett.* **127A** (1988) 112.

[11] R. E. Prange and S. M. Girvin (editors), *The Quantum Hall Effect* (Springer-Verlag, New York, 1987).

[12] R. Mackenzie and F. Wilczek, *Int. J. Mod. Phys.* **A3** (1988) 2827;
F. Wilczek and A. Zee, *Phys. Rev. Lett.* **51** (1983) 2250.

[13] W. J. Zakrzewski, *Low Dimensional Sigma Models* (Adam Hilger, Bristol, 1989).

[14] J. Eells and L. Lemaire, *Bull. Lond. Math. Soc.* **20** (1988) 385.

[15] N. S. Manton, *Phys. Lett.* **110B** (1982) 54.

[16] A. D'Adda, M. Lüscher and P. DiVecchia, *Nucl. Phys.* **B146** (1978) 63.

[17] V. L. Golo and A. M. Perelomov, *Phys. Lett.* **79B** (1978) 112.

[18] H. Eichenherr and M. Forger, *Nucl. Phys.* **B155** (1979) 381.

[19] A. M. Perelomov, *Phys. Rep.* **146** (1987) 135.

[20] G. H. Derrick, *J. Math. Phys.* **5** (1964) 1252.

[21] S. Coleman, *Nucl. Phys.* **B262** (1985) 263.

[22] A. M. Perelomov and M. C. Prati, *Nucl. Phys.* **B258** (1985) 647.

[23] C. Nash and S. Sen, *Topology and Geometry for Physicists* (Academic Press, London, 1983).

[24] R. Rajaraman, *Solitons and Instantons* (North-Holland, Amsterdam, 1982).

[25] E. B. Bogomolny, *Sov. J. Nucl. Phys.* **24** (1976) 449.

[26] A. M. Din and W. J. Zakrzewski, *Nucl. Phys.* **B174** (1980) 397.

[27] G. S. Adkins, C. R. Nappi and E. Witten, *Nucl. Phys.* **B228** (1983) 552.

[28] N. S. Manton and P. J. Ruback, *Phys. Lett.* **181B** (1986) 137.

[29] P. D. Lax, *Comm. Pure Appl. Math.* **21** (1968) 467.

[30] A. Ramani, B. Grammaticos and T. Bountis, *Phys. Rep.* **180** (1989) 159.

[31] V. E. Zakharov and A. V. Mikhailov, *Sov. Phys. JETP* **47** (1978) 1017.

[32] K. Pohlmeyer, *Comm. Math. Phys.* **46** (1976) 207.

[33] A. Davey and K. Stewartson, *Proc. Roy. Soc. Lond.* **A338** (1974) 101.

[34] D. Anker and N. C. Freeman, *Proc. Roy. Soc. Lond.* **A360** (1978) 529.

[35] S. Novikov, S. V. Manakov, L. P. Pitaevskii and V. E. Zakharov, *Theory of Solitons* (Consultants Bureau, New York, 1984).

[36] R. S. Ward, *J. Math. Phys.* **29** (1988) 386.

[37] R. S. Ward, *Nonlinearity* **1** (1988) 671.

[38] P. Forgács, Z. Horváth and L. Palla, *Nucl. Phys.* **B229** (1983) 77.

[39] N. C. Freeman, *Adv. App. Mech.* **20** (1980) 1.

[40] L. J. Mason and G. A. J. Sparling, *Phys. Lett.* **137A** (1989) 29.

[41] B. Berg and M. Lüscher, *Nucl. Phys.* **B190** (1981) 412.

[42] R. Ben-Av and S. Solomon, *Int. J. Mod. Phys.* **A5** (1990) 427.

[43] F. Fucito and S. Solomon, *Nucl. Phys.* **B251** (1985) 505.

[44] M. Göckeler, *Fortschr. Phys.* **36** (1988) 801.

[45] A. V. Mikhailov and A. I. Yaremchuk, *Nucl. Phys.* **B202** (1982) 508.

[46] R. S. Ward, *Phys. Lett.* **158B** (1985) 424.

[47] I. Stokoe and W. J. Zakrzewski, *Z. Phys.* **C34** (1987) 491.

[48] M. Peyrard and D. K. Campbell, *Physica* **9D** (1983) 33.

[49] J. J. M. Verbaarschot, T. S. Walhout, J. Wambach and H. W. Wyld, *Nucl. Phys.* **A461** (1987) 603.

[50] A. E. Allder, S. E. Koonin, R. Seki and H. M. Sommermann, *Phys. Rev. Lett.* **59** (1987) 2836.

[51] K. J. M. Moriarty, E. Myers and C. Rebbi, *Phys. Lett.* **207B** (1988) 411, *J. Comp. Phys.* **81** (1989) 481.

[52] E. P. S. Shellard and P. J. Ruback, *Phys. Lett.* **209B** (1988) 262.

[53] R. S. Ward, *Comm. Math. Phys.* **79** (1981) 317;
M. K. Prasad, *Comm. Math. Phys.* **80** (1981) 137;
E. Corrigan and P. Goddard, *Comm. Math. Phys.* **80** (1981) 575.

[54] P. J. Ruback, *Nucl. Phys.* **B296** (1988) 669.

[55] M. F. Atiyah and N. J. Hitchin, *Phys. Lett.* **107A** (1985) 21, *Phil. Trans. Roy. Soc. Lond.* **A315** (1985) 459.

[56] M. F. Atiyah and N. J. Hitchin, *The Geometry and Dynamics of Magnetic Monopoles* (Princeton University Press, Princeton, 1988).

[57] N. S. Manton and T. M. Samols, *Phys. Lett.* **215B** (1988) 559.

[58] A. M. Din and W. J. Zakrzewski, *Nucl. Phys.* **B253** (1985) 77.

[59] H. Jeffreys and B. Jeffreys, *Methods of Mathematical Physics, Third Edition*, §*25.12* (Cambridge University Press, Cambridge, 1956).

[60] W. J. Zakrzewski, in preparation.

[61] G. W. Gibbons and N. S. Manton, *Nucl. Phys.* **B274** (1986) 183.

[62] R. A. Leese, M. Peyrard and W. J. Zakrzewski, Durham preprint DTP-89/45 (1989), to appear in *Nonlinearity*.

[63] A. A. Bogolubskaya and I. L. Bogolubsky, *Phys. Lett.* **136A** (1989) 485.

[64] D. K. Campbell, J. F. Schonfeld and C. A. Wingate, *Physica* **9D** (1983) 1.

[65] D. K. Campbell and M. Peyrard, *Physica* **18D** (1986) 47.

[66] A. M. Safian, S. Coleman and M. Axenides, *Nucl. Phys.* **B297** (1988) 498;
A. M. Safian, *Nucl. Phys.* **B304** (1988) 392.

[67] J. A. Frieman and B. W. Lynn, *Nucl. Phys.* **B329** (1990) 1.

[68] R. Friedberg, T. D. Lee and A. Sirlin, *Phys. Rev.* **D13** (1976) 2739.

[69] R. Friedberg, T. D. Lee and A. Sirlin, *Nucl. Phys.* **B115** (1976) 1, 32.

[70] R. Friedberg and T. D. Lee, *Phys. Rev.* **D15** (1977) 1694.

[71] T. D. Lee, *Phys. Rev.* **D35** (1987) 3637;
R. Friedberg, T. D. Lee and Y. Pang, *Phys. Rev.* **D35** (1987) 3640, 3658;
T. D. Lee and Y. Pang, *Phys. Rev.* **D35** (1987) 3678.

[72] T. I. Belova and A. E. Kudryavtsev, *Sov. Phys. JETP* **68** (1989) 7;
Yu. A. Simonov and J. A. Tjon, *Ann. Phys.* **129** (1980) 110.

[73] Y. S. Kivshar and B. A. Malomed, *Rev. Mod. Phys.* **61** (1989) 763.

[74] R. S. Ward, *Phil. Trans. Roy. Soc. Lond.* **A315** (1985) 451;
A. N. Leznov and M. V. Saveliev, *Comm. Math. Phys.* **74** (1980) 111.

[75] M. K. Prasad, *Physica* **1D** (1980) 167.

[76] C. N. Yang, *Phys. Rev. Lett.* **38** (1977) 1377.